Developmental Testbed Center Report
AOP 2015 Activities

1 April 2015 – 31 March 2016

1 Introduction

The Developmental Testbed Center (DTC) is a distributed facility with components at the National Center for Atmospheric Research (NCAR) and the National Oceanic and Atmospheric Administration (NOAA) Earth System Research Laboratory (ESRL) Global Systems Division (GSD). The purpose of the DTC is to provide a link between the research and operational communities so results of research in Numerical Weather Prediction (NWP) can be efficiently transferred to operations. In addition, the DTC provides the research community access to a number of the latest operational NWP code packages for research applications. The DTC meets its goals by: maintaining and supporting community code packages that represent the latest NWP technology, performing extensive testing and evaluation (T&E) of new NWP technology, developing and maintaining a state-of-the-art verification package, and connecting the NWP research and operational communities through workshops and its visitor program. Over the past year, DTC activities were organized into six focus areas: Verification, Mesoscale Modeling, Data Assimilation, Hurricanes, Ensembles and Global Model Test Bed (GMTB). The GMTB is a new effort within the DTC that started in July 2015. This work is focused on facilitating the Research to Operations (R2O) process for the continued development of NOAA's Next Generation Global Prediction System (NGGPS).

Funding for the DTC is provided by NOAA's National Weather Service (NWS) and Office of Oceanic and Atmospheric Research (OAR), the Air Force (AF), NCAR, and the National Science Foundation (NSF). This report provides a description of the activities undertaken by the DTC between 1 April 2015 and 31 March 2016. These activities include those described in the DTC 2015 Annual Operating Plan (AOP), as well as a few carry-over activities from the DTC AOP 2013 and 2014.

1.1 DTC Management

The external management structure of the DTC includes an Executive Committee (EC), a Management Board (MB), and a Science Advisory Board (SAB). Current memberships are listed below. The MB and EC are responsible for approving the DTC AOP, which defines the work to be undertaken by the DTC in a given year, whereas the SAB is charged with providing the DTC Director with advice on future directions of the DTC and reviewing proposals submitted to the DTC Visitor Program. Over the past year, the DTC hosted its annual SAB meeting at NCAR's Foothills Campus in Boulder, CO, on 16-18 September 2015. The purpose of this meeting was to discuss strategic future directions for the DTC. The DTC also hosted a two-hour conference call with MB members on 6 October 2015 to report on recommendations from the SAB, discuss initial guidance on priorities for AOP 2016 and review status of the DTC Terms of Operation (TOO). On 27-28 January 2016, the DTC hosted its annual in-person MB meeting at NCAR's Foothills Campus in Boulder, CO. The purpose of this meeting was to discuss the DTC AOP 2016 and nominations for SAB members whose term will expire in June 2016. DTC management also participated in two DTC EC conference calls (25 June 2015, 20 November 2015) and an in-person EC meeting at NWS Headquarters in Silver Spring, MD, on 24 February 2016. Recent DTC accomplishments, recommendations from the SAB, proposed activities for AOP 2016, and the future direction of the DTC were discussed at the in-person meeting. The EC also approved the DTC Director's proposal to renew the term for one SAB member whose term expires in June 2016 (S. R. Gopalakrishnan), rotate off five

SAB members whose term expires in June 2016 (Evan Kuchera, Gary Lackmann, David Novak, Carolyn Reynolds and Robert J, Trapp) and add five new SAB members (term begins 1 July 2016). The five new SAB members are: Tom Auligne (Joint Center for Satellite Data Assimilation-JCSDA), Tim Whitcomb (Naval Research Laboratory-NRL), David Vollmer (United States AF Academy), Kathy Gilbert (NCEP's Weather Prediction Center-WPC) and Zhuo Wang (University of Illinois). Quarterly reports on the progress to date were also prepared for each activity and distributed to the EC and MB members.

DTC External Management Committees:

Executive Committee		Management Board			
Jim Hurrell	NCAR	Josh Hacker	NCAR	Hendrick Tolman	NOAA/NWS
Bill Lapenta	NOAA/NWS	Joe Klemp	NCAR	Fred Toepfer	NOAA/NWS
Ralph Stoffler	AF	Michael Gremillion	AF	Stan Benjamin	NOAA/OAR/ESRL
Kevin Kelleher	NOAA/OAR	John Zapotocny	AF	Tom Hamill	NOAA/OAR/ESRL

Science Advisory Board	
Adam Clark	National Severe Storms Laboratory
Robert Fovell	State University of New York (SUNY) – Albany
Kristen Corbosiero	SUNY – Albany
Sharanya Majumdar	University of Miami
David Novak	National Centers for Environmental Prediction (NCEP)/WPC
Geoff DiMego	NCEP/Environmental Modeling Center (EMC)
Jenni Evans	Pennsylvania State University
David Gochis	NCAR
S. R. Gopalakrishnan	NOAA/Atlantic Oceanographic and Meteorological Laboratory (AOML)
Evan Kuchera	AF
Gary Lackmann	North Carolina State University
Carolyn Reynolds	NRL
Brad Colman	Climate Corporation
Robert J. Trapp	Purdue University
Kelly Mahoney	Cooperative Institute for Research in Environmental Sciences
Russ Schumacher	Colorado State University
Kayo Ide	University of Maryland

1.2 Community Interactions

Maintaining strong ties to both the research and operational NWP communities is critical to the DTC's ability to successfully meet its mission. Over the past year, strong ties with the operational community were maintained through the DTC's interactions with our partners at the operational centers (i.e., EMC and AF) both at the management level and through our team lead interactions with the appropriate team leads and/or focal points at the operational centers. The DTC also worked toward strengthening its ties to the broader research community through workshops, tutorials and the DTC Visitor Program. Information on DTC-sponsored tutorials is provided in Section 2.3. The DTC also engages the community through the distribution of its newsletter "Transitions" that serves as a forum for the research and operational communities to share information. Over the past year, the DTC distributed three issues of Transitions. All issues of Transitions can be accessed at: http://www.dtcenter.org/newsletter/. In addition to these on-going efforts, the DTC continued to engage in discussions related to a NWP Information Technology Environment (NITE), which would facilitate the use of operational NWP systems by a broader spectrum of the research and development (R&D) community.

1.2.1 Community Outreach Events

In June 2015, the DTC co-hosted with NCAR's Mesoscale and Microscale Meteorology (MMM) Laboratory, the 16th WRF Users' Workshop at NCAR's Center Green Campus in Boulder, CO. The first day consisted of lectures on fundamentals of Weather Research and Forecasting (WRF) dynamics and best-practices, followed by a 3-day workshop consisting of 66 talks and approximately 70 posters. The last day consisted of six mini-tutorials on radar data assimilation, Big Weather, Hurricane WRF (HWRF), Mesoscale Model Evaluation Testbed (MMET), Visualization and Analysis Platform for Ocean, Atmosphere and Solar Researchers (VAPOR), and NCAR Command Language (NCL). The HWRF and MMET instructional sessions were organized and conducted by DTC staff. About 200 people attended the workshop (http://www2.mmm.ucar.edu/wrf/users/workshops/WS2015/WorkshopPapers.php).

In January 2016, the DTC hosted the Future of Statistical Post-processing in NOAA and the Weather Enterprise Workshop at the NOAA Center for Weather and Climate Prediction (NCWCP) in College Park, MD. The goal of this workshop was to help NOAA set its future requirements for providing internal and external customers with the high-quality data they need to achieve the expected benefits from statistical post-processing. The first two days of the workshop focused on collecting information on requirements, new developments in post-processing methods and community infrastructure, with the third day being comprised of productive breakout group discussions. Over 90 people from a broad spectrum of NOAA, international centers, NCAR, universities and the private sector participated in this workshop (http://www.dtcenter.org/events/workshops16/post-processing/).

In February 2016, the DTC hosted two back-to-back workshops related to sea ice modeling at NCAR's Center Green Campus in Boulder, CO. The Office of Naval Research (ONR) Sea State Modeling/Forecasting Workshop on 2 February focused on assessing the performance of prototype sea ice and wave modeling initiative led by NOAA/ESRL and NRL in support of the ONR Sea State Developmental Research Initiative field campaign. The NGGPS Sea Ice Modeling Workshop on 3-4 February focused on collecting input from the sea ice modeling community toward formulating a recommendation to NOAA's NGGPS program office on the choice of a sea ice model for inclusion in the NCEP Unified Global Coupled System (UGCS). The organizing committee for the NGGPS workshop represented a broad group of stakeholders in sea ice modeling community and was comprised of Marika Holland (NCAR), Janet Intrieri (NOAA/OAR/ESRL), Richard Allard (NRL), Cecilia Bitz (University of Washington), Robert Grumbine (NOAA/NWS), Annarita Mariotti (NOAA/OAR/Climate Program Office), and Eugene Petrescu (NOAA/NWS Alaska Region Headquarters). By scheduling the two workshops in sequence, the NGGPS Sea Ice Modeling Workshop benefitted from a larger community presence and from the lessons learned about short-term, high-resolution, sea ice forecasts in configurations both coupled and uncoupled with the atmosphere. These workshops, which included over 50 scientists representing both the research and operational communities, brought together experts from the community to discuss the state of sea ice and coupled modeling systems, forecasting, and predictability to inform the selection of the UGCS sea ice model. Discussions also focused on enhancing collaborations and highlighting areas of needed R&D. More information on the workshops, as well as all of the presentations, can be found on the workshop website: http://www.dtcenter.org/events/workshops16/seaice/. The report for the NGGPS workshop is undergoing final review and will be posted on the workshop website in April 2016.

1.2.2 DTC Visitor Program

The DTC Visitor Program supports visitors to work with the DTC to test new forecasting and verification techniques, models and model components for NWP. The goal is to provide the operational weather prediction centers (e.g., NCEP and AF) with options for near-term advances in operational weather forecasting and to provide researchers with NWP codes that represent the latest advances in technology. It also offers an opportunity for visitors to introduce new techniques that would be of particular interest to the research community into the publicly-released software systems supported by the DTC.

Over the past year, the DTC received a final project report from Dr. Fovell. Hence, all projects selected in January 2013 are now complete (see Table 1.2.2-1). In addition to working towards wrapping up projects selected in 2013, the DTC provided support for six projects selected for funding in 2014 (see Table 1.2.2-2). Dr. Yablonsky completed the ocean model portion of his project. The deliverables for his project included a report, as well as new software capabilities that were transitioned to the DTC. Yablonsky's project also included delivering a wave model capability, work that was completed under the direction of Dr. Isaac Ginis due to Dr. Yablonsky leaving the University of Rhode Island (URI) to take another job. An HWRF three-wave air-sea-wave coupled framework was made available to EMC's HWRF developers through this visitor project and a combined report covering both aspects of the project was delivered to the DTC. Dr. Wang completed his ensemble project over the summer. In addition to his project report, the new ensemble capability at the center of Dr. Wang's project was delivered to the GSD node of the DTC. Dr. Geresdi submitted his final project report in the fall. Dr. Bao completed his development of new satellite diagnostic capabilities that have been incorporated into the DTC tool set and provided a project report describing the outcome of applying his tools to HWRF retrospective forecasts. Dr. Galarneau delivered his report on diagnostics of the 2015 HWRF Retrospective Test. All visitor project reports received over the past year are available on the "Visitor Program" portion of the DTC website (http://www.dtcenter.org/visitors/). Dr. Roebber's project is nearing completion. Four more proposals were selected for funding in 2015 (see Table 1.2.2-3) and two more projects were awarded funding in early 2016 (see Table 1.2.2-4). All 2015 projects are progressing nicely.

Table 1.2.2-1. 2013 Visitor Projects

PI	Institution	Project Title
Robert Fovell / Peggy Bu (graduate student)	University of California – Los Angeles	Improving HWRF Track and Intensity Forecasts Via Model Physics Evaluation and Tuning

Table 1.2.2-2. 2014 Visitor Projects

PI	Institution	Project Title
Shaowu Bao	Coastal Carolina University	Evaluation of Two HWRF Microphysics/Radiation Configurations with Remote-Sensing Data
István Geresdi	University of Péc	Towards Improving Representation of Convection and MCC Longevity in High-Resolution WRF and NEMS-NMMB Model Forecasts
Hongli Wang	Colorado State University	Estimation of Initial and Forecast Error Variances for the NCEP's Operational Short-Range Ensemble Forecast (SREF) System
Richard Yablonsky	URI	Developing and Supporting Global HWRF Ocean Coupling with Advanced Ocean Physics and Initialization Options and New Diagnostic Tools for Comprehensive Model Evaluation
Thomas Galarneau	NCAR	Diagnosing Tropical Cyclone Motion Forecast Errors in the 2014 HWRF Retrospective Test (H214)
Paul Roebber	University of Wisconsin-Milwaukee	Demonstration Project: Development of a Large Member Ensemble Forecast System for Heavy Rainfall using Evolutionary Programming

Table 1.2.2-3. 2015 Visitor Projects

PI	Institution	Project Title
Jason Otkin	University of Wisconsin - Madison	Object based verification for the HRRR model using simulated and observed GOES infrared brightness temperatures
Gretchen Mullendore	University of North Dakota	Mesoscale Model Intercomparison at Convection-Allowing Resolution using MODE
Dev Niyogi	Purdue University	Improving WRF Weather Forecast through Enhanced Representation of Cropland-Atmosphere Interactions
Joel Bedard	University of Quebec - Montreal	Implementation and validation of a geo-statistical observation operator for the assimilation of near-surface winds in GSI

Table 1.2.2-4. 2016 Visitor Projects

PI	Institution	Project Title
Michael Iacono	Atmospheric and Environmental Research	Testing revisions to RRTMG cloud radiative transfer and performance in HWRF
Robert Fovell	SUNY-Albany	Impact of planetary boundary layer assumptions on HWRF

1.2.3 NWP Information Technology Environment (NITE)

For scientists outside of the NWS to contribute relevant R&D to NCEP's modeling suite, it is important for them to work with the current operational codes, workflows, and input datasets. However, obtaining such codes and inputs, and configuring the system to run with data assimilation and cycling workflows identical to those used in operations, can be a daunting task for the research community. For AOP 2014, the DTC undertook the task of assembling a preliminary design for a framework referred to as NITE, that would facilitate preparing and running research experiments using NCEP's modeling systems. Over the past year, the NITE team continued to publicize its final report describing the NITE system design, and prepared a manuscript that was submitted to the Bulletin of the American Meteorological Society (BAMS). NITE also has substantial overlap with the goals of the NWS's NGGPS program, so the DTC is actively participating in several NGGPS teams who are addressing similar system framework issues.

2 Software Systems

To serve as a bridge between operations and research, the DTC provides a framework for the two communities to collaborate in order to accelerate the transition of new scientific techniques into operational weather forecasting. This framework is based on software systems that are a shared resource with distributed development. The current operational systems are a subset of the capabilities contained in these software systems. Ongoing development of these systems is maintained under version control with mutually agreed upon software management plans. The DTC currently works with the following software systems:

- WRF – NWP model + pre- and post-processors
- HWRF - set of tools for tropical storm forecasting, including a coupled atmosphere and ocean system
- NOAA Environmental Modeling System (NEMS) / Nonhydrostatic Multiscale Model on the B grid (NMMB) – NWP model + pre-processor
- Unified Post-Processor (UPP)

- Gridpoint Statistical Interpolation (GSI) data assimilation (DA) system
- Ensemble Kalman Filter (EnKF) DA System
- Modular end-to-end ensemble system
- Model Evaluation Tools (MET) – Verification package

The DTC does not generally contribute to the development of new scientific techniques for these software packages. The two exceptions are MET development and some limited physics package development for WRF to address short-comings brought to light by DTC T&E. The DTC contributes to the software management of all of these systems and user support for the publicly-released systems (WRF, NMMB, HWRF, UPP, GFDL vortex tracker, GSI, EnKF and MET). All software management and user support activities are collaborative efforts with the developers, where the exact role of the DTC depends on the software package. The main developers of these packages are affiliated with EMC, ESRL, NCAR, Global Modeling and Assimilation Office (GMAO) of the National Aeronautics and Space Administration (NASA), National Environmental Satellite, Data and Information Service (NESDIS), Geophysical Fluid Dynamics Laboratory (GFDL), URI and the Hurricane Research Division (HRD) of NOAA's AOML. DTC activities are currently focused on the regional application of these software systems. With the recent addition of the GMTB, this focus may broaden in the future, but for AOP 2015 the focus for the packages mentioned above remained regional. In addition to working with these individual software systems, the DTC is involved in efforts to develop and maintain scripting and workflows for a number of forecast systems: HWRF, Rapid Refresh (RAP), North American Mesoscale Rapid Refresh (NAMRR), High Resolution Ensemble Forecast (HREF) system and the Global Forecast System (GFS). These workflows provide an important framework for conducting carefully controlled T&E activities.

For the GMTB, the DTC is working with EMC and the community to establish a Common Community Physics Package (CCPP) that will serve as a framework for efficiently transitioning the development of next generation physics parameterizations into operations to meet the needs of NWS's NGGPS. Another important component of this work is establishing an Interoperable Physics Driver (IPD), which will provide a framework for physical parameterization suites within the CCPP to properly interface with different dynamic cores. Establishing a code management plan and user support for these packages will build on the DTC's extensive experience working with the community to establish frameworks to facilitate distributed development. Over the past nine months, the GMTB team assembled a requirements document for the IPD/CCPP package. These requirements were presented to EMC and the National Unified Operational Prediction Capability (NUOPC) Physics Interoperability group and the document was updated to incorporate feedback from these groups. These requirements are serving as a foundation for discussions with EMC directed toward developing plans for making this concept a reality. In addition to the requirements document, the DTC prepared draft coding standards and code management documents for IPD/CCPP package that were sent to EMC for review. Another important component of a framework for facilitating distributed development is documentation. The DTC has begun compiling technical documentation for the current version of the IPD and the physics suite used for the deterministic application of NCEP's GFS. In an effort to reduce the cost of maintaining documentation for a package whose content will evolve over time, the DTC explored different options for generating documentation from source code and decided to use Doxygen. Using this approach, the DTC has generated a template for individual physical parameterization documentation and produced complete documentation for the GFS planetary boundary layer (PBL) and radiation schemes, with the remaining members of the operational physics suite soon to follow. The updated code that includes new in-line comments capable of generating web-based documentation has been sent to EMC for review.

2.1 Software Management

While specific software management plans differ between the various software packages, they all contain the following elements:

- Code repositories maintained under version control software.
- Protocols for proposing modifications to the software, whether the modifications are simply updates to current features, bug fixes or the addition of new features.
- Testing standards proposed software modifications must pass prior to being committed to the code repository.
- Additional testing standards used to more thoroughly check the integrity of the evolving code base.

Given all these software packages continue to evolve over time, all testing standards must be updated periodically in order to meet the maintenance requirements of the code base. Over the past year, the DTC continued to collaborate with the various developer groups on these ongoing software management activities. The DTC also continued to provide a pathway for the research community to contribute to the development of these software systems. Noteworthy events from this work over the past year are:

- ***WRF*** –Work is underway to incorporate code changes related to the inclusion of a new smooth terrain-following hybrid-vertical coordinate for the Advanced Research WRF (ARW) dynamic core while minimizing impacts on other parts of the modeling system. All 2015 operational HWRF forecast system capabilities were committed to the WRF trunk, and were made available for the next community release of WRF. These capabilities include the Ferrier-Aligo high-resolution microphysics scheme [available for both ARW and Nonhydrostatic Mesoscale Model on the E grid (NMME)], updates to the surface flux exchanges with the coupled ocean (NMME only), implementation of aerosol-aware Rapid Radiative Transfer Model for General Circulation Models (RRTMG) and Thompson schemes for NMME, updates for nesting and feedbacks (NMME only), as well as miscellaneous bug fixes and changes to tuning parameters. Changes to the WRF code to improve the efficiency of nest parallelization and quilting for NMME were also tested and transitioned to the WRF repository. Additionally, the DTC worked closely with MMM to prepare for the WRF repository transition from Subversion to Git in the summer of 2016.

- ***UPP*** – The DTC continued to work closely with EMC to manage the UPP code base through regular bi-monthly meetings. Efforts to keep the community UPP repository in sync with EMC's operational UPP repository are ongoing. The most recent community release of UPP included GRIB2 output capability along with full NMMB functionality and support. The DTC also established a new stand-alone UPP webpage, Users' Guide, and email help desk (upp-help@ucar.edu).

- ***NEMS*** –The DTC continued to enhance the portability of the NEMS software package and associated libraries. Work to enhance the efficiency of the model through domain decomposition was undertaken. The Thompson aerosol-aware microphysics scheme was integrated into NMMB by DTC staff and provided to EMC for inclusion in the NEMS repository, which will provide the opportunity for future T&E using this new option.

- ***HWRF*** – The DTC continued to support HWRF developers in using and adding innovations to the code repository. Development of Python scripts for integration of the multistorm capability in the HWRF code base was completed and committed to the HWRF repository. Changes to the scripts to improve nest parallelization and quilt efficiency were tested and transitioned to the

HWRF code repository. All 2015 operational HWRF forecast system capabilities were committed to the HWRF code repository and made available to the community in the last HWRF release. These capabilities included upgrades to WRF, GSI, UPP and the vortex relocation, as well as extending atmospheric-ocean coupling to all basins and enhanced product generation. In addition to software maintenance activities, the DTC provided enhanced support for Hurricane Forecast Improvement Project (HFIP)-funded principal investigators contributing to HWRF development. The DTC provided coordination of development activities by chairing the HWRF developers committee bi-weekly meetings and hosting two HWRF specific Python training sessions to support active developers. These training sessions, held in conjunction with the HFIP Annual Meeting and the January HWRF Tutorial, consisted of four hours of lectures followed by hands-on activities. Approximately 25 developers attended each session either in-person or remotely. Materials and resources from the developer training, including a video of the second session, are available at: http://www.dtcenter.org/HurrWRF/developers/ docs/documents.php. To facilitate inter-developer collaboration, DTC began hosting an *hwrf-contrib* repository for peer-to-peer sharing of code. The HWRF helpdesk was transitioned to Request Tracker (RT). This new ticketing system is used for tracking both support requests from HWRF developers working directly with the repository and users working with the public release code. The HWRF v3.7a Scientific Documentation was published as an NCAR technical note:

Tallapragada, V., L. Bernardet, M. K. Biswas, I. Ginis, Y. Kwon, Q. Liu, T. Marchok, D. Sheinin, B. Thomas, M. Tong, S. Trahan, W. Wang, R. Yablonsky, X. Zhang, 2016: Hurricane Weather Research and Forecasting (HWRF) Model: 2015 Scientific Documentation. NCAR Technical Note NCAR/522+STR, 116 pp.

The HWRF Users' Guide was submitted for publication as a NOAA GSD Technical Memorandum:

Biswas, M. R., L. Carson, C. Holt, and L. Bernardet. Community HWRF Users' Guide v3.7a. NOAA GSD Technical Memorandum. Submitted, 152 pp.

- ***GSI and EnKF*** – The DTC continued to perform code reviews for each proposed code update and synchronize the DTC community code repository with the trunk of EMC's operational repository. Through the community code repository, the DTC continued to provide repository code access to HWRF community developers and public releases and support to general developers and users. The DTC continued to facilitate the joint GSI and EnKF Review Committee and hosted an onsite meeting at NCAR's Foothills Campus on 10 August 2015. The DTC continued its effort to unify the compilation utility used by EMC and DTC for GSI, EnKF, and NCEP libraries.

2.2 Verification Tool Development

The DTC verification team completed development for the MET v5.1 and released the code to the community on 26 October 2015. The release included six new tools, support for three new datasets, four new ways of handling point data, eight major enhancements, twenty minor enhancements, and twenty-four new measures added to the MET output. Several of these new capabilities were added to MET through projects outside the DTC, but the regression testing and documentation prior to the release and helpdesk support after the release were provided by the DTC. Documentation for MET was migrated to an improved framework to allow the entire development team to work on documentation at one time rather than one person at a time. New plot types, Performance Diagrams and Probability Integral Transform Histogram (PHIST), user-interface enhancements, and bug fixes were added to METViewer, as well as the additional measures added to MET output.

The new tools include regrid_data_plane and shift_data_plane to manipulate Grib1, Grib2 and Network Common Data Form (NetCDF) files prior to matching forecast and analysis/observation fields. MET v5.1 also includes tools to read the innovation (O-B) and analysis increment (O-A) from GSI binary files and compute matched pairs from both single-value solutions (gsid2mpr) and ensemble-based solutions (gsiens2orank). Once reformatted, the Stat-Analysis tool can be used to perform additional filtering or calculate statistics. The gen_vx_mask tool replaces the gen_poly_mask and gen_circle_mask tools and adds support for track masking and accumulating multiple masks together. Finally, the Method for Object-based Diagnostic Evaluation (MODE) Time Domain (MTD) tool was added to operate on time-series of gridded data fields to define three-dimensional space-time objects. An example of MTD objects is provided in Fig. 2.2-1. Objects were computed from 25% probability of accumulated snow exceeding 3 inches over 6 hours. Twelve forecasts valid on 23 January 2016 at 15 UTC were used for this evaluation. The view is from northeast of the United States. The colors indicate object movement from west (blue) to east (red-orange).

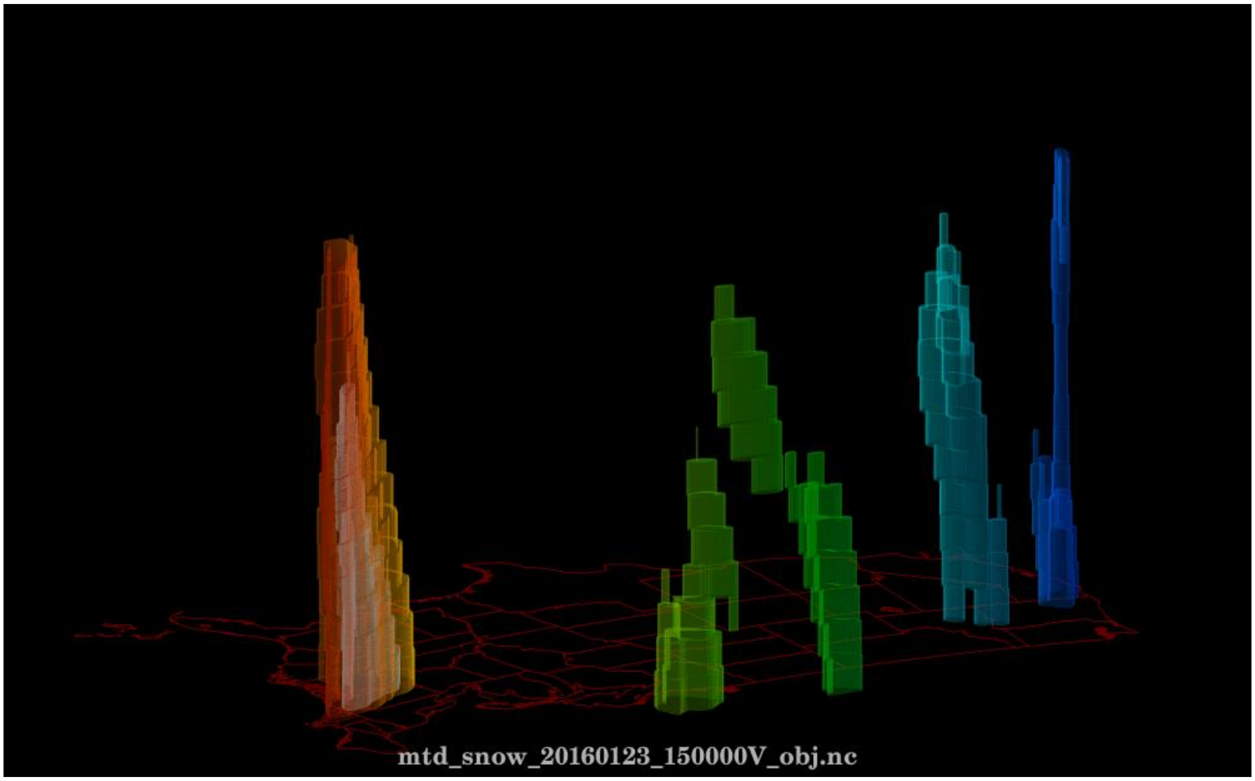


Figure 2.2-1. Example of MODE-Time Domain objects computed from 25% probability of accumulated snow exceeding 3 inches over 6 hours.

In collaboration with the NEMS Global Aerosol Component group within EMC's Global Weather and Climate Modeling Branch (GCWMB), support to read AErosol Robotic NETwork (AERONET) data was added to MET's Ascii2NC preprocessing tool and provides the capability to process any high frequency data into user-defined averages or quantiles. Support was added to the MADIS2NC tool for the Western Wind and Solar Integration Study (WWSIS) dataset and Aircraft Communications Addressing and Reporting System (ACARS) profiles.

A major enhancement to METv5.1 is the ability to regrid gridded fields within the Point-Stat, Grid-Stat, Ensemble-Stat, Series-Analysis, MODE and MTD tools. The functionality is identical to that in regrid_data_plane but is performed in memory, which streamlines the verification process and significantly reduces storage requirements. This approach also makes the use of climatologies for computation of skill scores more feasible. Another major enhancement in this MET release is the inclusion of conditional thresholding prior to computing continuous statistics, which allows measures such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE) to be computed over user defined ranges using simple thresholding (>, <, =) operators. Complex thresholding was also included to allow the user to compound several simple thresholds together using & (and) and II (or) operators. The compound operator concept was also added to the mask generating tool (gen_vx_mask) and allows the user to define complex regions to evaluate scores over intersections, subsets or unions of two or more masking regions. Figure 2.2-2 shows an example of generating a “water-only” mask for superstorm Sandy (22 October – 2 November 2012). The land mask (upper left) and a mask of 200 km around the track of superstorm Sandy in 2012 (upper right) were combined using the command “Not Land && Sandy” to produce the resulting mask (bottom). Other enhancements include added flexibility in TC-Stat for defining tropical cyclone rapid-intensification and rapid-weakening events and generalized for use with other fields (e.g. renewable energy ramp events) in Stat-Analysis. The swinging door algorithm for detecting change of intensity (e.g. ramp) events was also added.

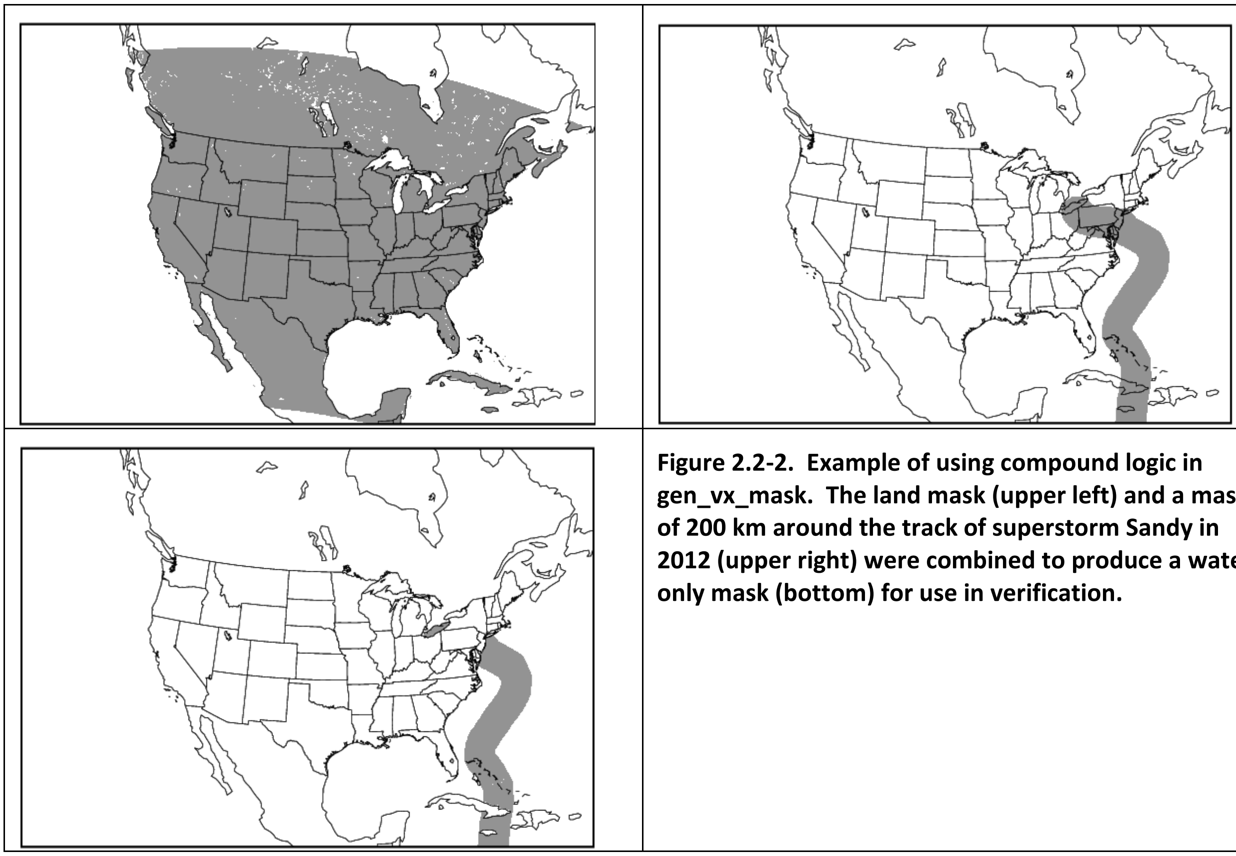


Figure 2.2-2. Example of using compound logic in gen_vx_mask. The land mask (upper left) and a mask of 200 km around the track of superstorm Sandy in 2012 (upper right) were combined to produce a water-only mask (bottom) for use in verification.

During AOP2015, the DTC worked with EMC and GSD to document the capabilities needed to facilitate unification of verification between the three organizations. This preliminary inventory was summarized in spreadsheet and shared with the NGGPS Verification and Validation Team. In response to requirements specified by EMC’s GCWMB, Point-Stat, Grid-Stat, Ensemble-Stat and Series-Analysis tools were enhanced to read in a climatological (or reference) mean or probability field to support computing Anomaly Correlation and skill scores such as Brier Skill Score and Continuous Ranked Probability Skill Score. Through this work, the need to add scalar anomaly and vector anomaly line types to Grid-Stat and the Series-Analysis tools was identified as well. Note that these new scores are based on standard methods and do not yet exactly emulate the methods used by NCEP. The DTC verification team is working with contacts at EMC to understand their methods to facilitate the inclusion of these methods in the next minor MET release. In response to requests from MET-TC users at the National Hurricane Center (NHC), the ability to ingest sixteen additional storm properties from the operational model output files (referred to as adecks) and NHC best track analysis (referred to as bdeck) was added to MET-TC. These properties include: pressure in millibars of the last closed isobar (RADP), radius of the last closed isobar in nautical miles (RRP), radius of maximum wind in nautical miles (MRD), wind gusts in knots (GUSTS), eye diameter in nautical miles (EYE), direction of storms in compass coordinates (DIR), storm speed in knots (SPEED), and system depth classification (DEPTH).

The DTC verification team has also been working with the AF 557th Weather Wing to explore the use of MET and the R-statistics package for cloud verification. The goal is to develop a set of metrics that can be used by the AF to assess the quality of their cloud forecasts. The AF 557th will assess how to and if blending the measures into an NWP index will be useful. This activity did not get underway until September 2015 due to reprogramming of AF funding for the DTC. Initial advancements were limited by challenges associated with transferring sample data between AF and DTC. However, MTD and several R-statistics methods were demonstrated on a sample case and have provided ideas for beneficial enhancements to MET during the coming year. For example, the computation of distance maps and the measures that are derived from them (e.g. the Baddely Delta metric). Completion of the AOP 2015 cloud verification work is anticipated in June 2016.

2.3 Publicly-Released Systems

The DTC currently collaborates with developers on eight software systems that undergo a public release process: WRF, UPP, HWRF, NEMS/NMMB, GFDL vortex tracker, GSI, EnKF and MET. Assistance continued to be offered through email helpdesks for all packages. Information regarding the timing and version of the most recent release, along with the current number of registered users and average helpdesk tickets per month for each package are listed in Table 2.3-1. Table 2.3-2 contains a list of the web addresses for each software package’s users’ page.

In addition to general MET user support, the DTC verification team actively responded to requests from NOAA users regarding the use of MET and METViewer. Accomplishments over the past year include: 1) increased registered MET users from NOAA (from 35 to 74); 2) worked with NCEP Central Operations (NCO) to install a development version of METViewer for use by EMC staff; 3) installed METViewer at GSD for evaluation by the High Impact Weather Prediction Program (HIWPP); 4) trained staff in EMC’s Mesoscale Modeling Branch (MMB) and GCWMB and GSD on the use of METViewer, 5) supported GSD’s Global Observing Systems Analysis (GOSA) group and GFDL’s staff on understanding the capability and use of MET; and 6) participated in the NGGPS Verification and Validation Team discussions. The DTC verification team has also been working with NCO to increase storage and troubleshoot the METViewer instance on EMC’s development server. Through its collaboration with AF 557th Weather Wing, support for the AF team on use of MET has also increased from one to two helpdesk tickets submitted per year to three to five helpdesk tickets per month at the end of AOP 2015.

The DTC hosted the Joint Community GSI and EnKF Data Assimilation System Tutorial at NCAR’s Foothills Laboratory in Boulder, Colorado on 11-14 August 2015. This event marked the sixth community tutorial for GSI but the first community tutorial for EnKF. The combined tutorial was a four-day venture with lectures given by system developers along with practical hands-on sessions. The tutorial reached a maximum capacity with 41 students from the U. S. and the international community. All slides from the tutorial are available at http://www.dtcenter.org/com-GSI/users/docs/index.php (under 2015 Community GSI/EnKF Tutorial).

The DTC hosted a community HWRF tutorial at the NCWCP in College Park, MD on 25-27 January 2016. The three-day tutorial included lectures covering all aspects of the HWRF system, given by system developers from EMC, HRD, URI, and DTC. The tutorial included 27 participants affiliated with universities, private companies, and research laboratories. In addition to the January HWRF tutorial, the DTC was invited to organize an HWRF tutorial at Nanjing University of Information Science and Technology in Nanjing, China. The tutorial exceeded capacity, attracting 84 registered participants. The DTC's contributions for this event included: overall planning, preparation and delivery of lectures and hands-on practical sessions and coordination of temporary accounts on NCAR's Yellowstone for tutorial participants. Materials from both events are posted at http://www.dtcenter.org/HurrWRF/users/tutorial/index.php.

Table 2.3-1: Code releases, number of registered users and number of helpdesk tickets per month for the publicly-released software packages supported by the DTC over the past year.

Software Package	Public Release			
	Version	Timing	Registered Users	Helpdesk tickets per month
WRF	V3.7	May 2015	~28,600	~400
	V3.7.1	August 2015		
UPP	V3.0	May 2015	398	~10
NEMS	v1.0	March 2016	150	<5
HWRF	V3.7a	August 2015	1205	~30
GFDL Vortex Tracker	V3.5a	September 2013	544	
GSI	V3.4	July 2015	1,573	~30
EnKF	V1.0	July 2015		~10
MET	V5.1	October 2015	3020	~15-20

Table 2.3-2: Users page websites for publicly-released software packages.

Software Package	Users Websites
WRF	http://www.mmm.ucar.edu/wrf/users/
UPP	http://www.dtcenter.org/upp/users/
NEMS	http://www.dtcenter.org/nems-nmmb/users/
HWRF	http://www.dtcenter.org/HurrWRF/users/
GSI	http://www.dtcenter.org/com-GSI/users/
EnKF	http://www.dtcenter.org/EnKF/users/
MET	http://www.dtcenter.org/met/users/

3 Testing and Evaluation

T&E activities undertaken by the developers of new NWP techniques from the research community are generally focused on case studies. However, in order to adequately assess these new technologies, extensive T&E must be performed to ensure they are indeed ready for operational consideration. DTC T&E generally focuses on extended retrospective time periods. The cases selected incorporate a broad range of weather regimes ranging from null, to weak and strong events. The exact periods chosen vary based on the phenomenon of focus for the test. The technique to be tested must be part of the code repositories supported by the DTC to ensure that the code has reached a certain level of maturity. The

DTC's evaluation of these retrospective forecasts includes standard verification techniques, as well as new verification techniques when appropriate. All verification statistics undergo a statistical significance (SS) assessment when appropriate. By conducting carefully controlled, rigorous testing, including the generation of objective verification statistics, the DTC is able to provide the operational community with guidance for selecting new NWP technologies with potential value for operational implementation. DTC testing also provides the research community with baselines against which the impacts of new techniques can be evaluated. The statistical results may also aid researchers in selecting model configurations to use for their projects.

3.1 Mesoscale Modeling

Mesoscale NWP systems are utilized in both research and operational forecasting applications and can be configured to suit a broad spectrum of weather regimes. Due to the number of approaches developed and offered by NWP systems, it is necessary to rigorously test select configurations and evaluate their performance for specific applications.

3.1.1 Testing Protocol and MMET

MMET (http://www.dtcenter.org/eval/meso_mod/mmet) provides the opportunity for the research community to conduct their own T&E of a new technique. Datasets for sixteen cases, deemed to be of high interest by EMC, are distributed via RAMADDA, a Repository for Archiving, Managing and Accessing Diverse DAta (http://ramadda.org/). Over the past year, three new cases were added to MMET. MMET datasets include a variety of initialization and observation datasets, as well as baselines for select operational configurations that were established by the DTC utilizing the MMET datasets.

DTC staff collaborated with scientists from NOAA ESRL's Physical Sciences Division (PSD) to identify an atmospheric river (AR) event (13-16 February 2011) from the Hydrometeorology Testbed (HMT) to include in MMET. The DTC's contribution to this effort was funded under its carry-over US Weather Research Program (USWRP) funding from AOP 2013. This particular MMET case was run for a 12-km Contiguous United States (CONUS) parent domain and a 3-km nest centered over California (Fig. 3.1.1-1) with model output every 15 minutes. In addition to the baseline results provided for select operational configurations, a second ARW configuration was run for the AR event using a physics suite defined and tuned by PSD staff. Plots of integrated vapor transport, as well as the scripts used to produce them (contributed by HMT staff), are also provided for this AR event through RAMADDA.

The scope of MMET was also broadened by adding two hurricane cases associated with Hurricane Edouard. Both cases (12th and 15th of September 2014) provide a unique opportunity to evaluate HWRF forecasts using extensive observations that were collected when NOAA deployed four Coyote Unmanned Aircraft Systems. The 12 September 2014 case provides a good example of a right of track bias for an HWRF forecast, with weaker than observed intensity. Conversely, the HWRF intensity forecast for the 15 September 2014 case exhibited large fluctuations during the first 12 hours, which may be related to initialization issues. In addition to providing the necessary input datasets for running the full HWRF system, users also have the flexibility to utilize input files to just run WRF coupled to the Message Passing Interface Princeton Ocean Model for Tropical Cyclones (MPIPOM-TC) and perform sensitivity experiments without running all the HWRF components. In addition to the GRIB outputs from all three HWRF domains, verification output of track and intensity is provided, along with links to additional observations such as dropsondes, Stepped-Frequency Microwave Radiometer, and radar datasets for users who want to evaluate HWRF. Plotting tools are made available to plot track, intensity, and other 3D fields.

In addition to the new cases added to MMET, maintenance of existing cases was conducted. This started with updating all software components to the most recent released versions. Forecasts were then generated for eight select cases using the most recent version of the respective model code base. Several new enhancements to the MMET workflow were also implemented over the past year, including: 1) nesting for both WRF-ARW and NEMS-NMMB, 2) ability to re-grid observations to the post-processed forecast domain within the grid-to-grid verification step (i.e., as opposed to pre-processed outside the workflow), 3) a new WRF-ARW baseline using RAP/HRRR operational physics suites, and 4) 6-hr warm-start capability using the GSI DA package (currently only functional in WRF-ARW).

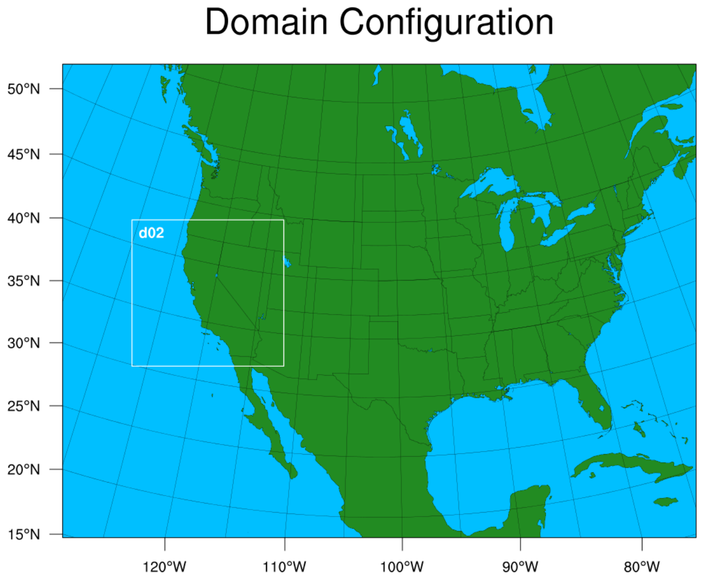


Figure 3.1.1-1. Computational domains used for the 13-16 February 2011 AR case added to MMET. The outer box defines the 12-km CONUS domain (d01), while the inner white box defines the 3-km nest (d02).

An update on MMET was presented at both the 16th WRF Users' Workshop (June 2015 - http://www2.mmm.ucar.edu/wrf/users/workshops/WS2015/ppts/6a.7.pdf) and American Meteorological Society (AMS) NWP / Weather Analysis and Forecasting (WAF) conference (July 2015 - https://ams.confex.com/ams/27WAF23NWP/webprogram/Paper273501.html) to publicize and demonstrate how the community can make use of this resource. During the final day of the 16th WRF Users' Workshop, a 1.5 h MMET instructional session was offered for approximately 25 participants (http://www2.mmm.ucar.edu/wrf/users/workshops/WS2015/ppts/MMET.pdf). Community interaction activities also included hosting a visitor from the Chinese Meteorological Administration to work with MMET using their Global/Regional Assimilation and Prediction System model. To continue to broaden the research community's knowledge of MMET, a manuscript was submitted to and accepted by BAMS:

Wolff, J. K., M. Harrold, T. Hertneky, E. Aligo, J. Carley, B. Ferrier, G. DiMego, L. Nance, Y.-H. Kuo, 2016: Mesoscale Model Evaluation Testbed (MMET): A resource for transitioning NWP innovations from research to operations (R2O). *Bull. Amer. Meteor. Soc.* In press.

3.1.2 NEMS

3.1.2.1 Impact of Thompson/RRTM on NAM physics suite

A comprehensive T&E activity employing the NEMS-NMMB model infrastructure was finalized, with the focus of the test on assessing the impact of replacing the Ferrier-hires microphysics scheme with the

Thompson microphysics scheme within NMMB for the NAM application (Table 3.1.2.1-1). This T&E activity, which was a carry-over activity from AOP 2013, employed a parent domain at 12-km grid spacing and two nests at 3-km grid spacing: CONUS and Alaska (Fig. 3.1.2.1-1). The end-to-end system included the NMMB Preprocessing System (NPS), NMMB, UPP, and MET software packages. The NEMS-NMMB code used for this T&E activity corresponded to the Friendly User Release v0.9. The testing period included one month per season in 2013-2014 (Table 3.1.2.1-2) with cases initialized every 36 h and run out to 48 h (for a total of 94 potential cases), providing a distribution of both 00 and 12 UTC initializations.

Table 3.1.2.1-1. Physics suites for the NMMB T&E activity.

Physics Scheme	Baseline Configuration (NAMOC)	Replacement Configuration (ThompsonMP)
Microphysics	Ferrier-hires	Thompson
Radiation Shortwave (SW) and Longwave (LW)	Rapid Radiative Transfer Model (RRTM)	Rapid Radiative Transfer Model (RRTM)
Surface Layer	Mellor-Yamada-Janjic	Mellor-Yamada-Janjic
Land Surface Model	Noah	Noah
PBL	Mellor-Yamada-Janjic	Mellor-Yamada-Janjic
Convection	Betts-Miller-Janjic (parent only)	Betts-Miller-Janjic (parent only)

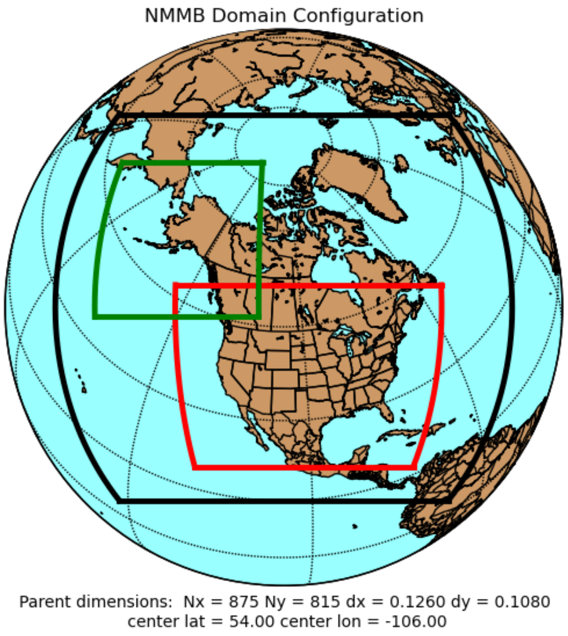


Figure 3.1.2.1-1. Computation domains used for the NMMB T&E activity. The black box defines the 12-km parent domain, the red box defines the 3-km CONUS nest, and the green box defines the 3-km Alaska nest.

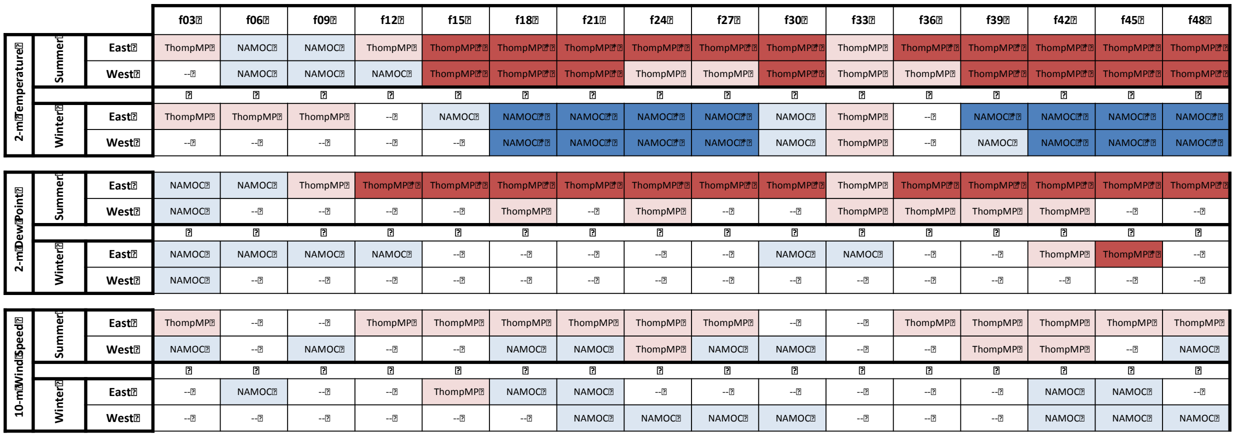
Table 3.1.2.1-2. Dates for the NMMB T&E activity.

Season	Dates
Fall	12 Oct – 15 Nov 2013
Winter	16 Jan – 19 Feb 2014
Spring	16 Apr – 17 May 2014
Summer	6 Jul – 9 Aug 2014

The full evaluation included an assessment of several different variables. In terms of traditional verification approaches, the surface and upper-air temperature, dew point temperature, and wind speed were evaluated using bias-corrected root mean squared error (BCRMSE) and bias. The precipitation accumulation and composite reflectivity variables were evaluated using Gilbert Skill Score (GSS) and frequency bias. For each of the evaluated parameters, confidence intervals (CIs) at the 99% level were applied to objectively assess SS and practical significance (PS). Further analysis was conducted to compare model output fields from each configuration, including SW and LW radiation, surface fluxes (sensible and latent), and PBL height. The final report and extensive supporting information can be found on the testing and evaluation portion of the DTC website (http://www.dtcenter.org/eval/meso_mod/nmmb_test/nems_v0.9/index.php). A brief summary highlighting a few key results for the CONUS 3-km nest are presented here.

Overall, when looking at the surface variables, a large number of SS and PS pair-wise differences were observed; however, which configuration was favored depended on verification metric, temporal aggregation, initialization time, vertical level, lead time, and threshold (Table 3.1.2.1-3). For 2-m temperature (Fig. 3.1.2.1-2), a notable result was that both configurations exhibited warm biases during the summer that grew with increased forecast lead time; however, an opposite signal was seen in the winter aggregation, where there were cold biases during the daytime hours. When differences were present, ThompsonMP typically had colder median bias values than NAMOC, leading to better performance by ThompsonMP in the summer when there was a warm bias and better performance by NAMOC in the winter when cold biases were present. Spatial distribution plots helped diagnose regional patterns that may not be captured in the time series plots. When examining 2-m temperature, all seasons showed NAMOC having a higher mean bias across the CONUS with the East having higher mean biases compared to the West. During the summer, a predominantly warm bias existed over the CONUS, while the winter displayed a cold bias over much of the CONUS, which is noticeably enhanced for the ThompsonMP configuration. When looking at 2-m dew point temperature (not shown), both configurations showed dry biases increasing with forecast lead time during the summer and moist biases during the winter with ThompsonMP generally the better performer. While no PS pair-wise differences were noted for 10-m wind speed bias, notable regional results showed the West CONUS typically having a neutral-to-low bias while the East CONUS had a consistent high bias regardless of season (not shown).

Table 3.1.2.1-3. Statistically significant (light shading) and practically significant (dark shading) differences for 2-m temperature, 2-m dew point temperature, and 10-m wind speed bias by season, region, and forecast lead time for the 00 UTC initializations of the CONUS 3-km nest.



			f03	f06	f09	f12	f15	f18	f21	f24	f27	f30	f33	f36	f39	f42	f45	f48
2-m Temperature	Summer	East	ThompMP	NAMOC	NAMOC	ThompMP	ThompMP*	ThompMP*	ThompMP*	ThompMP*	ThompMP*	ThompMP*	ThompMP	ThompMP*	ThompMP*	ThompMP*	ThompMP*	ThompMP*
		West	--	NAMOC	NAMOC	NAMOC	ThompMP*	ThompMP*	ThompMP*	ThompMP	ThompMP	ThompMP*	ThompMP	ThompMP	ThompMP*	ThompMP*	ThompMP*	ThompMP*
	Winter	East	ThompMP	ThompMP	ThompMP	--	NAMOC	NAMOC*	NAMOC*	NAMOC*	NAMOC*	NAMOC	ThompMP	--	NAMOC*	NAMOC*	NAMOC*	NAMOC*
		West	--	--	--	--	--	NAMOC*	NAMOC*	NAMOC*	NAMOC*	NAMOC	ThompMP	--	NAMOC	NAMOC*	NAMOC*	NAMOC*
2-m Dew Point	Summer	East	NAMOC	NAMOC	ThompMP	ThompMP*	ThompMP*	ThompMP*	ThompMP*	ThompMP*	ThompMP*	ThompMP*	ThompMP	ThompMP*	ThompMP*	ThompMP*	ThompMP*	ThompMP*
		West	NAMOC	--	--	--	--	ThompMP	--	ThompMP	--	--	ThompMP	ThompMP	ThompMP	ThompMP	--	--
	Winter	East	NAMOC	NAMOC	NAMOC	NAMOC	--	--	--	--	--	NAMOC	NAMOC	--	--	ThompMP	ThompMP*	--
		West	NAMOC	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
10-m Wind Speed	Summer	East	ThompMP	--	--	ThompMP	ThompMP	ThompMP	ThompMP	ThompMP	ThompMP	--	--	ThompMP	ThompMP	ThompMP	ThompMP	ThompMP
		West	NAMOC	--	NAMOC	--	--	NAMOC	NAMOC	ThompMP	NAMOC	NAMOC	--	--	ThompMP	ThompMP	--	NAMOC
	Winter	East	--	NAMOC	--	--	ThompMP	NAMOC	NAMOC	--	--	--	--	--	--	NAMOC	NAMOC	--
		West	--	--	--	--	--	--	NAMOC	NAMOC	NAMOC	NAMOC	--	--	--	NAMOC	NAMOC	NAMOC

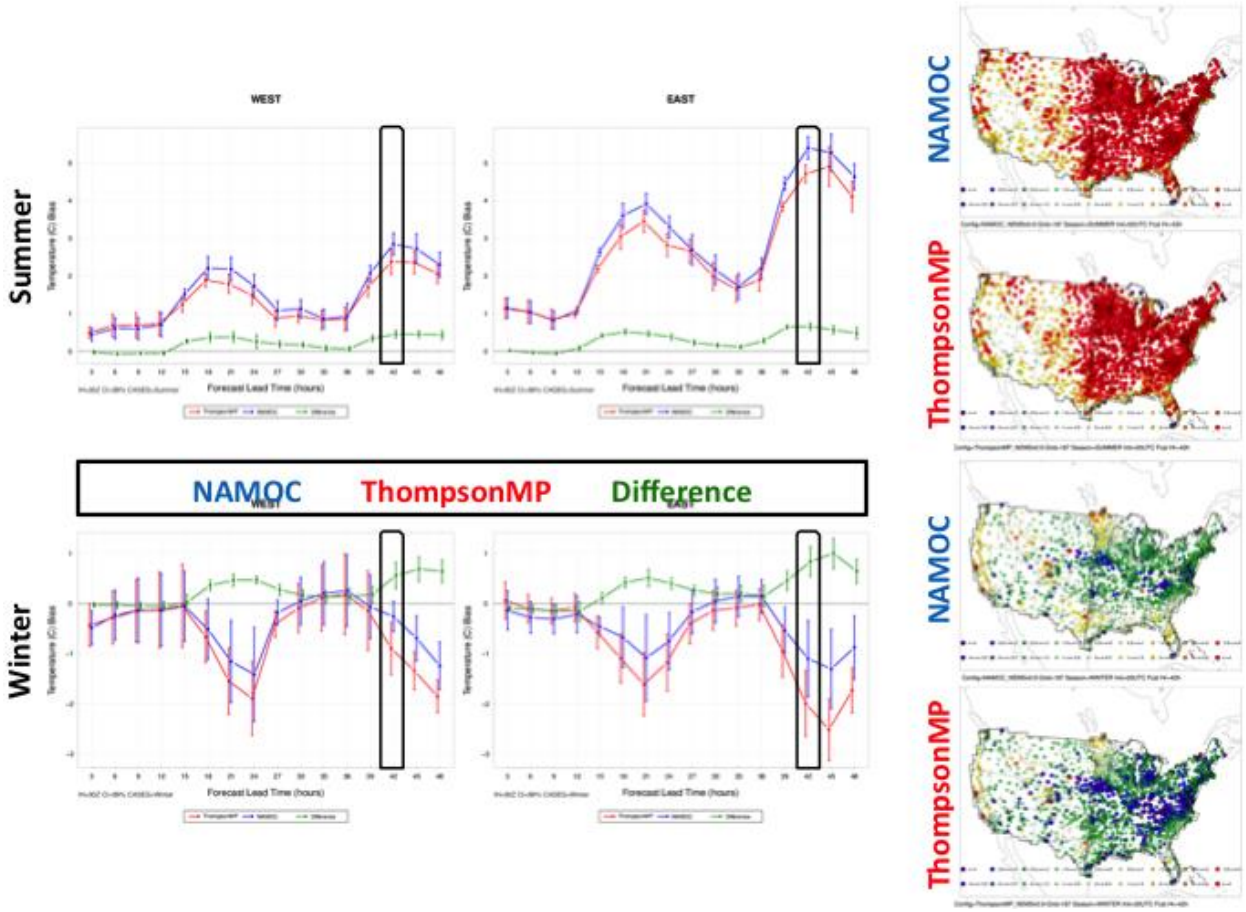


Figure 3.1.2.1-2. Time series of 2-m AGL temperature (°C) median bias for all 00 UTC initializations for the CONUS West (left) and CONUS East (center) domains for the summer (top) and winter (bottom) aggregations. NAMOC is in blue, ThompsonMP in red, and the differences (NAMOC-ThompsonMP) in green. The vertical bars attached to the median represent the 99% CIs. Spatial plots (right column) of the 2-m AGL temperature (°C) mean bias for the 42-h forecast lead time for all 00 UTC initializations for the summer (top two panels) and winter (bottom two panels) aggregations.

An examination of daily precipitation accumulation over the CONUS showed minimal differences between the two configurations with few differences being SS (not shown). On the other hand, an examination of composite reflectivity showed NAMOC having consistently higher frequency bias for all aggregations and forecast lead times for thresholds of 10 and 20 dBZ (Fig. 3.1.2.1-3). A number of SS differences were found, all favoring ThompsonMP.

A key result in this sensitivity study was the differences in SW and LW radiation reaching the surface between the two configurations. Given the ThompsonMP scheme is coupled with the RRTM scheme and passes the cloud water droplet, cloud ice and snow size distributions to the radiation scheme, impacts on both cloud-free and cloudy skies were expected. In general, this difference in microphysics schemes led to NAMOC having higher radiative values than ThompsonMP with more SW radiation reaching the surface (Fig. 3.1.2.1-4), resulting in higher upward LW radiation and higher near-surface temperatures CONUS-wide for all seasons.

The results obtained during the extended DTC testing of the NAMOC and ThompsonMP configurations were utilized by EMC along with internal results produced during their parallel runs to drive an evidence-based decision to remove the lower limit for cloud droplet effective radius in RRTM coupled with the Ferrier-Aligo microphysics. This modification is expected to reduce incoming surface SW radiation fluxes under liquid clouds and, in turn, reduce surface temperature warm biases. In addition, a partial cloudiness scheme to better represent subgrid scale clouds is also being implemented to further improve the surface temperature forecasts. Provided positive results are seen during EMC's pre-implementation testing, both of these modifications will be fully implemented in the next NAM operational bundle upgrade.

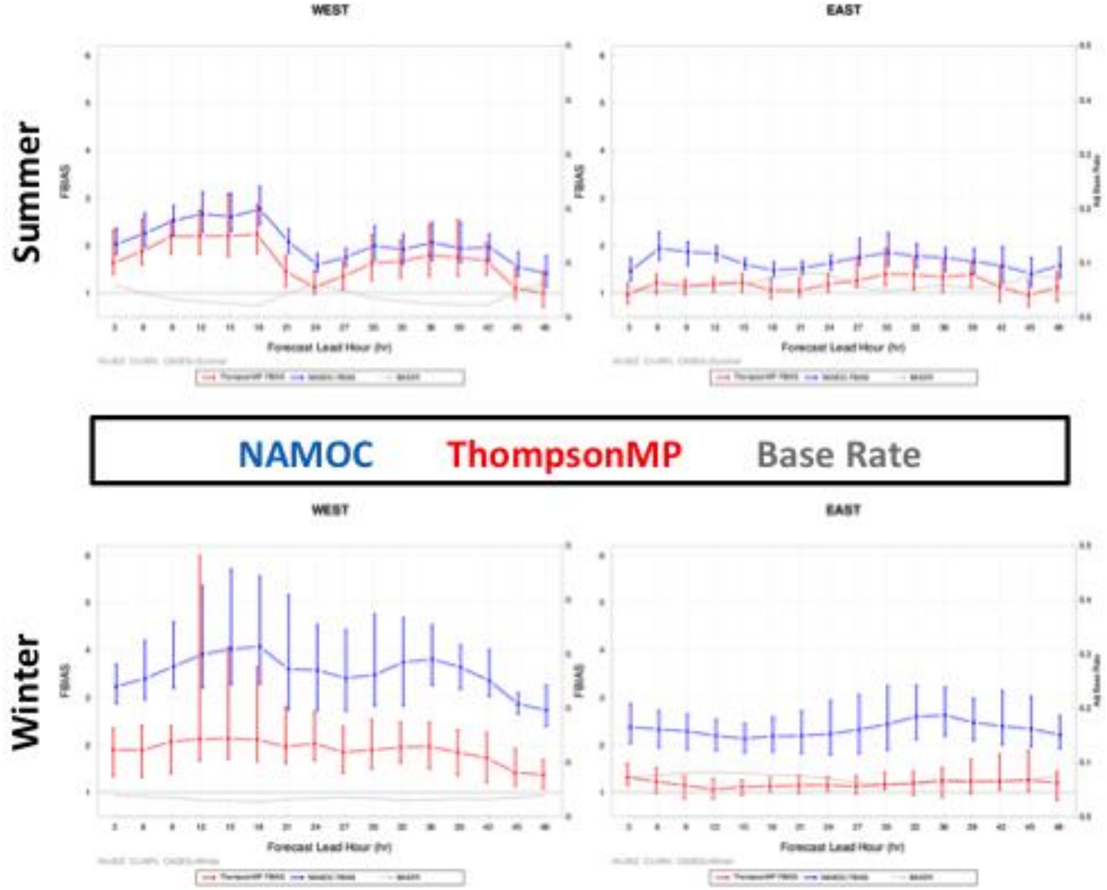


Figure 3.1.2.1-3. Time series of composite reflectivity (dBZ) frequency bias for the West CONUS (left) and East CONUS (right) for all 00 UTC initializations for reflectivity greater than 10 dBZ for the summer (top) and winter (bottom) aggregations. NAMOC is in blue and ThompsonMP in red. The base rate, in grey, is associated with the y-axis on the right. The vertical bars attached to the aggregated values represent the 99% CIs.

3.1.2.2 NAMRR T&E environment

Through collaborations with EMC, work was conducted to establish a T&E environment on the NCAR supercomputer, Yellowstone, using the Rocoto Workflow Management System that is functionally similar to the NAMRR system under development at EMC (Fig. 3.1.2.2-1). Code for the hourly updating NAMRR system was checked out of the EMC repository and ported to Yellowstone. The hourly end-to-end system using the NAM operational physics suite was successfully completed for the 12-km North American parent with two 3-km nested domains over the CONUS and Alaska. Example plots from one particular run of the NAMRR system on Yellowstone are shown in Fig. 3.1.2.2-2. DTC staff also worked to incorporate the community released versions of UPP and MET into the NAMRR workflow. For UPP, both the unipost and copygb functions were included to de-stagger fields, generate derived meteorological variables, vertically interpolate fields to isobaric levels, and interpolate to operational grids (grids 218 and 242, and a user defined CONUSnest grid). Additional enhancements now use wgrib2 in place of copygb. The inclusion of MET in the NAMRR workflow focused on adding the necessary tools to perform grid-to-point and grid-to-grid verification tasks. In addition, a reformatting tool was included in the automated workflow to process the point observations into the expected format and interpolate the gridded observation datasets to a consistent grid (used in the UPP/wgrib2 step) for comparison purposes. Code enhancements necessary to run the system on Yellowstone, as well as those required for inclusion of the MET tasks, were ultimately merged back into the EMC repository. A presentation on the NAMRR T&E environment implemented at the DTC was given at the 16th WRF Users' Workshop (June 2015 - http://www2.mmm.ucar.edu/wrf/users/workshops/WS2015/posters/p34.pdf).

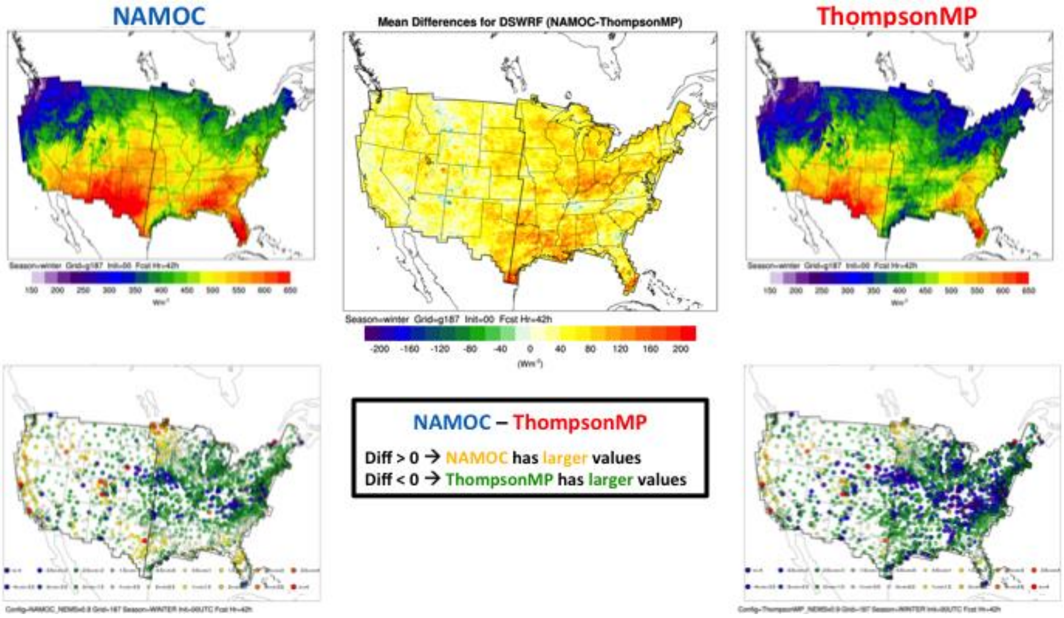


Figure 3.1.2.1-4. Mean downward short-wave radiation (W m^{-2}) for all 00 UTC initializations for the winter aggregation at the 42-h forecast lead time for the NAMOC configuration (top left), ThompsonMP configuration (top right), and mean differences (NAMOC-ThompsonMP; top center). Spatial plots of 2-m AGL temperature (°C) mean bias at the 42-h forecast lead time for all 00 UTC initializations for the winter aggregation of NAMOC (bottom left) and ThompsonMP (bottom right).

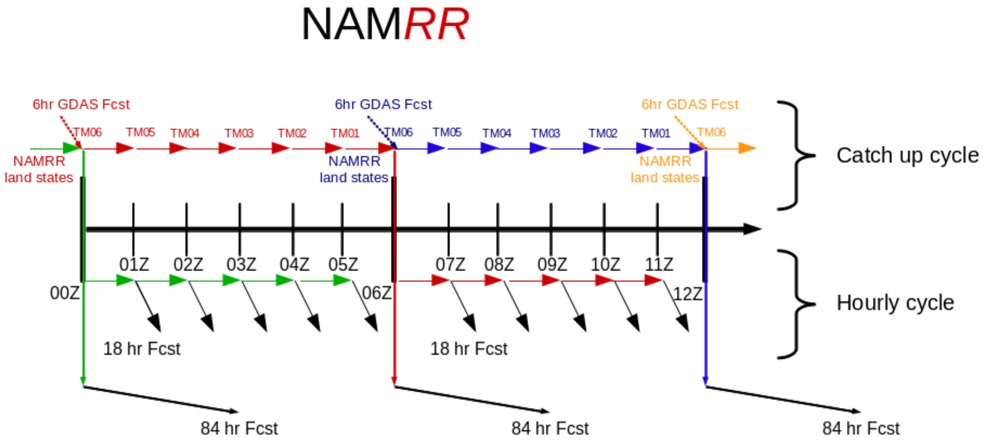


Figure 3.1.2.2-1. Schematic of the NAMRR flow diagram. Courtesy of Jacob Carley, EMC.

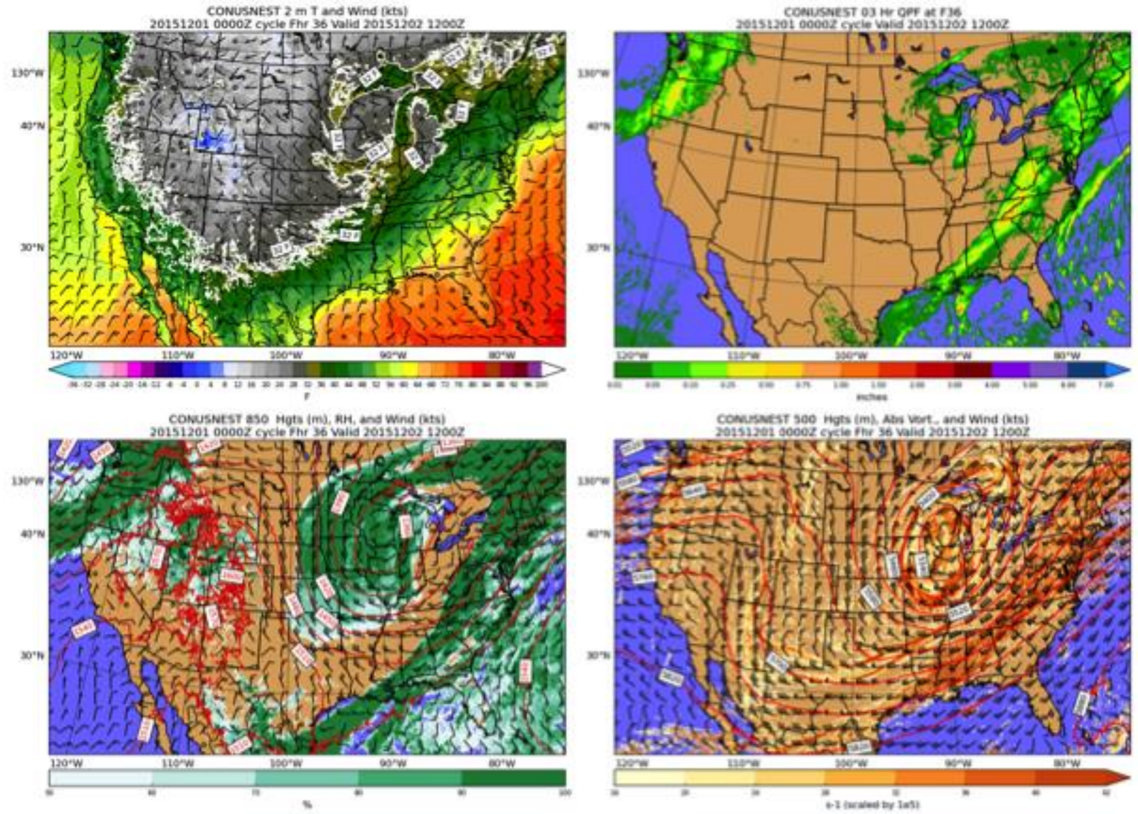


Figure 3.1.2.2-2. Example plots created by the hourly NAMRR system run on Yellowstone, including 2-m temperature and winds (top left), 3-h QPF (top right), 850-mb geopotential heights, relative humidity, and winds (lower left), and 500-mb geopotential heights, absolute vorticity, and winds (lower right).

3.1.2.3 Test of expanded HRRR-ARW domain

The HRRR model represents a major step forward in the operational prediction of severe thunderstorms and mesoscale convective systems as well as other year-round mesoscale phenomena. In addition to the original purpose to improve prediction of warm–season convection and its impact on the National Aerospace System, the HRRR has found wide acceptance by forecasters inside and outside of the NWS as guidance for a variety of weather phenomena in all seasons, including East Coast winter storms, winter precipitation type, timing and intensity of heavy non-convective and convective precipitation and land-falling tropical cyclones, and hub-height wind trends for the renewable-energy industry. As computing resources at NCEP continue to increase in coming years, we foresee that the current regional models, the NAM and RAP, will be replaced by regional cloud-resolving configurations of domain size similar to that of the current NAM and RAP. These cloud-resolving configurations will be nested within the future operational global model. Looking toward that day, and in view of the importance of accurate short-term forecasts for vulnerable coastal areas, particularly along the Gulf and Atlantic coasts, we proposed to investigate the value of an initial expansion of the HRRR domain in all directions, but mainly toward the east and south.

For the purpose of this activity two cases of interest were selected. The first case is 3-4 October 2015, a flooding event associated with hurricane Joaquin. The second case is 22-23 January 2016, a winter storm that affected the Mid-Atlantic states. All necessary data for these two simulations were obtained and staged on the NOAA super computer. The extended HRRR domain is presented in Figure 3.1.2.3-1.

The original plan was to initialize the expanded HRRR domain based on the full domain RAP forecasts on the rotated lat-long grid, but DTC staff discovered the current WRF Preprocessing System cannot read the ARW rotated lat-long grid. Since WPS is unable to read rotated lat-long grids, alternative ways of initializing HRRR are currently being explored.

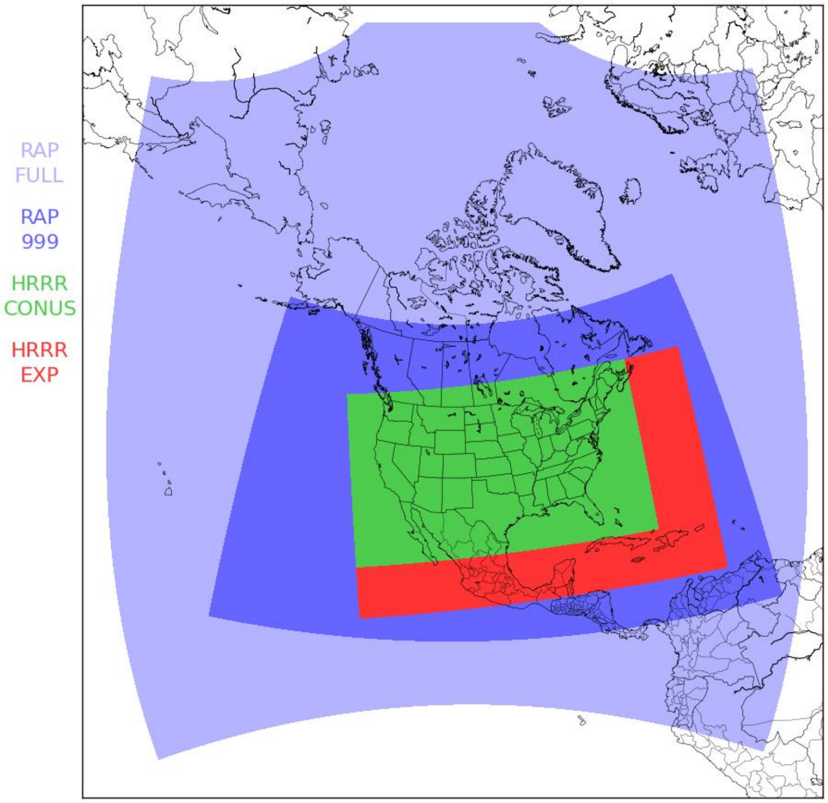


Figure 3.1.2.3-1. Full RAP domain (purple), old RAP domain (blue), original HRRR CONUS domain (green) and expanded HRRRR domain (red).

3.2 Hurricanes

3.2.1 HWRF Quantitative Precipitation Forecasts

A comprehensive evaluation of HWRF Quantitative Precipitation Forecasts (QPF) was conducted for the 2014 operational HWRF system. Precipitation forecast verification for the HWRF system has previously been limited, prompting this study to provide a baseline for future improvement and indicate where to focus development. The output used for this evaluation, which was provided by EMC, consisted of pre-implementation runs for 22 storms from 2011 to 2013 Hurricane seasons and a subset of operational forecasts from the 2014 season. Three basic approaches were applied to gain insight into the performance of HWRF QPF: 1) a large-scale assessment of HWRF QPF accumulated over 24-h for the parent domain, 2) 24-h accumulations for a circular region with a 600-km diameter centered on the observed storm location (with and without corrections for track forecast error), and 3) run-total storm-scale QPF for the innermost domain with 3-km grid spacing. The Climate Prediction Center's MORPHing technique (CMORPH) analyses and NCEP's Stage IV analyses (available only over the CONUS region) were used as sources of quantitative precipitation estimates (QPE) for these assessments. Preliminary results from this evaluation were presented at the 16th WRF Users' Workshop (June 2015) (http://www2.mmm.ucar.edu/wrf/users/workshops/WS2015/ppts/6a.1.pdf) and the AMS NWP /WAF

Conference (July 2015) (https://ams.confex.com/ams/27WAF23NWP/webprogram/Paper273615.html). Key results from this evaluation were summarized in the previous annual report. Preparation of a final report is underway as the results are being consolidated in a concise and informative manner. Schematics (as shown in Figure 3.2.1-1) were generated for better description of the methodology, in preparation for a manuscript.

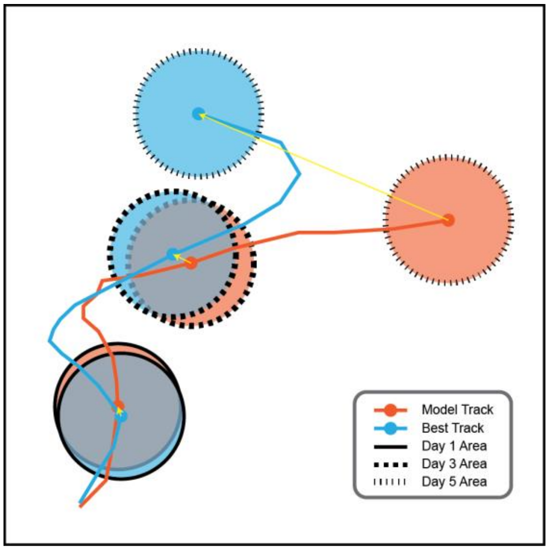


Figure 3.2.1-1. Schematic illustrating the horizontal shifting of the predicted precipitation grid to the location of the observed storm. The entire field is shifted to collocate the forecasted storm location (red) with the best track storm location (black) for each valid time (corresponding circles).

3.2.2 Rapid Intensification Forecasts

Rapid intensification (RI) events, which are defined as an intensity increase of 30 kt or more over-water in 24 h (Kaplan and Demaria 2003), are rare and difficult to predict. For this study, evaluations were done using retrospective forecasts for the Atlantic (AL) and eastern North Pacific (EP) basins from the 2014 Stream 1.5 exercise (sample includes storms from 2011-2013 Hurricane seasons), as well as real-time forecasts during the 2014 Hurricane season. Given the higher frequency of RI events in the western North Pacific (WP) basin, HWRF's ability to capture RI events was also evaluated for operational forecasts in the WP basin for the 2013 and 2014 seasons (sample includes two different model versions). Additionally, an evaluation was conducted using 2015 HWRF pre-implementation tests produced by EMC.

For the AL and EP basins combined, the 2014 HWRF system tended to under-predict the magnitude of the intensity change when it correctly forecasted a RI event. HWRF over-predicted intensity change by 15-20 kt for false alarms and under-predicted by 20-25 kt for missed RI events. For the WP basin, the outcome for missed events and false alarms was similar to that for the AL and EP basins. In contrast, the magnitude of the intensity change for the WP basin was not biased for correctly forecasted RI events. A homogeneous comparison of the 2014 and 2015 HWRF systems indicated the 2015 HWRF model was better at capturing intensity change (Table 3.2.2-1). Results from this evaluation were presented at the 23rd Conference on Probability and Statistics in the Atmospheric Sciences during the 96th AMS annual meeting (January 2016 - https://ams.confex.com/ams/96Annual/webprogram/Paper290802.html). The full report for the RI/RW evaluation is available on the DTC webpage: http://www.dtcenter.org/eval/hwrf_rirw/.

Table 3.2.2-1. Contingency tables for homogeneous comparison of HWRF 2014 model (top) and HWRF 2015 model (bottom) with observed and forecast events of an intensity increase of 30 kt or greater in 24-h for all lead times combined. Aggregation includes storms in the AL and EP basins.

		Model Forecast		
		RI	**No RI**	**Total**
Observation	**RI**	**42** (0.15%)	**1003** (3.55%)	**1045** (3.70%)
	No RI	**138** (0.49%)	**27037** (95.81%)	**27175** (96.3%)
	Total	**180** (0.64%)	**28040** (99.36%)	**28220** (100%)

		Model Forecast		
		RI	**No RI**	**Total**
Observation	**RI**	**72** (0.26%)	**973** (3.45%)	**1045** (3.70%)
	No RI	**169** (0.60%)	**27006** (95.70%)	**27175** (96.3%)
	Total	**241** (0.85%)	**27979** (99.15%)	**28220** (100%)

3.2.3 Advancing the Connections between Radiation and Clouds in HWRF

2013 T&E activities revealed that Thompson microphysics in the 2013 version of HWRF produced improvements in track for the AL basin, with degradations in track forecasts for the EP basin. Given the significant upgrades to the operational HWRF system after the 2013 version, performance when using the Thompson microphysics scheme within the 2015 HWRF was re-evaluated. This T&E activity was designed in close collaboration with the EMC HWRF team to inform 2016 pre-implementation testing, where the Thompson and advected Ferrier-Aligo microphysics schemes were both candidates for replacement of the non-advected Ferrier-Aligo microphysics scheme. The tests included five storms from the AL basin and eleven storms in the EP basin that occurred during the 2014 and 2015 seasons. Particular emphasis was placed on EP basin storms in response to the 2013 T&E results. The focus of the DTC's evaluation was the impact of replacing the non-advected Ferrier-Aligo microphysics scheme with the Thompson microphysics scheme. Prior to conducting the retrospective test, the Thompson microphysics scheme and the partial cloudiness scheme within the RRTMG radiation scheme were modified in an effort to understand the cause of, and then alleviate, the increased track error in the EP basin. These modifications included a bug fix to the partial cloudiness scheme within the radiation parameterization, fall speed changes within the Thompson microphysics scheme, and alterations to the RRTMG partial cloudiness scheme to change the lower limit of the snow and ice particle size.

The updated codes were initially tested on a well-understood case study of Hurricane Daniel in the EP basin. Results indicated promising improvements to the along- and cross-track errors. Track and intensity error statistics for the large test indicated the Thompson configuration produced smaller track errors than the Ferrier-Aligo configuration in the AL basin beyond 60 hours (Figure 3.2.3-1), and had no statistically significant intensity differences (not shown). Despite the modifications, the Thompson configuration still produced larger track errors than the Ferrier-Aligo configuration beyond 96 hours in the EP basin (Figure 3.2.3-1). These errors are dominated by along-track errors indicating the Thompson microphysics configuration moved storms too quickly (not shown). Although the sample for the current test is vastly different from that for the 2013 test, the magnitude of the track error differences between the Thompson configuration and the operational baseline are noticeably reduced from the 2013 T&E results beyond 72 hours.

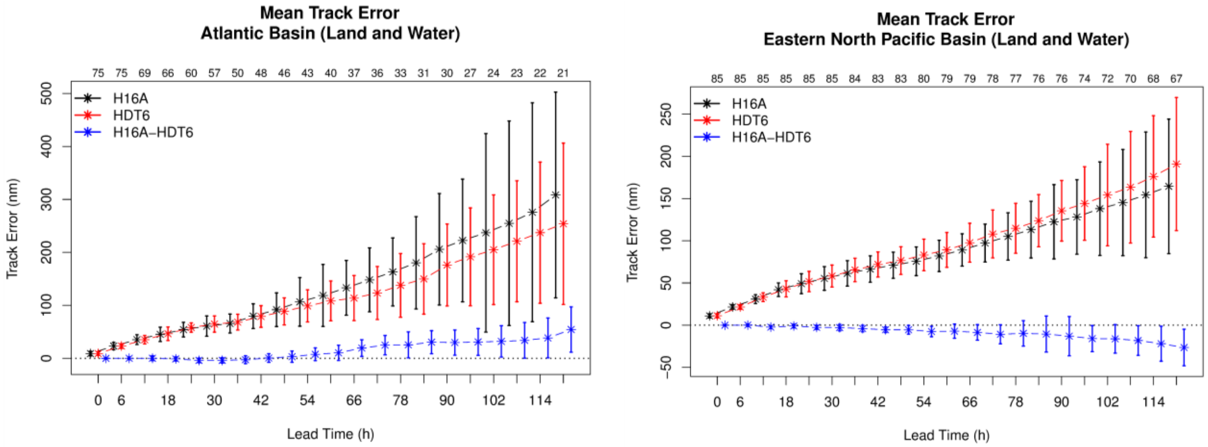


Figure 3.2.3-1. Mean track errors in the AL basin (left) and EP basin (right) with respect to lead time. Ferrier-Aligo microphysics is shown in black, Thompson microphysics in red, and mean pairwise differences (blue) with 95% CIs.

Intensity errors in the EP basin showed a large negative intensity bias for both configurations, with larger errors present in the Thompson microphysics configuration beyond 36 hours (Figure 3.2.3-2). The large intensity errors present in both configurations stem from the sample containing challenging intensity forecasts, including multiple RI cases.

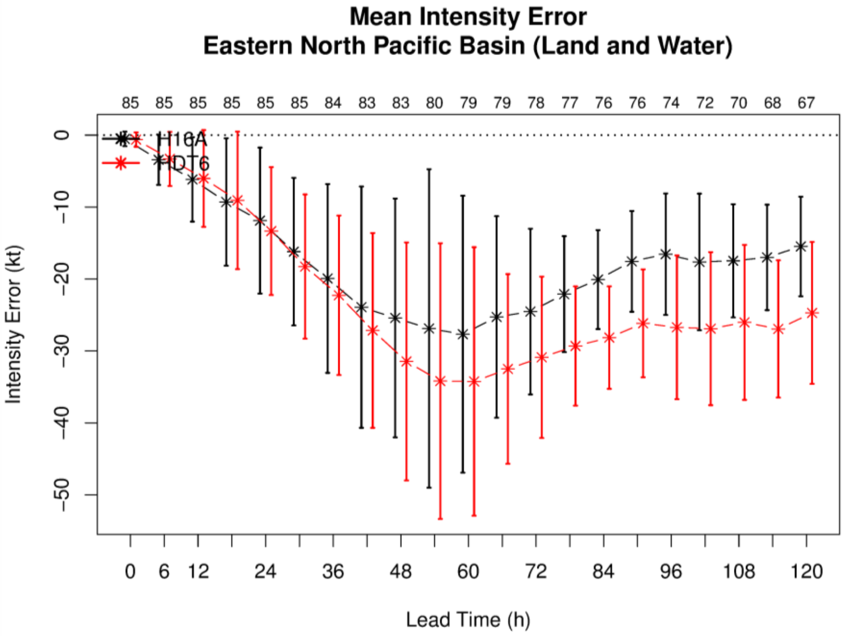


Figure 3.2.3-2. Same as 3.2.3-1, except mean intensity errors in the EP basin.

In addition to looking at track and intensity errors, the DTC evaluated the large-scale flow of the two configurations with a focus on the EP basin to better understand the differences in performance. GFS analyses were used to represent truth for this portion of the evaluation. For longer lead times, both configurations exhibited a westerly wind bias between 15°N and 0°, relative to the GFS analyses. The configuration using Thompson microphysics produced an environment with too much shear relative to the GFS analyses, whereas the operational version using the Ferrier-Aligo microphysics created an environment more similar to that of the GFS analysis (not shown). The difference plots in Fig. 3.2.3-3 indicate the Thompson configuration produced stronger westerly flow aloft (250 hPa) and less westerly flow at lower levels (850 hPa). The increased shear in the Thompson microphysics configuration is consistent with producing weaker storms, although the explanation for why the Thompson configuration produced more shear is under investigation. Both configurations had a warm and moist bias compared to GFS analyses. However, the large-scale differences reveal the Thompson configuration produced cooler temperatures and lower relative humidity in the mid-levels relative to the operational configuration, indicating an improvement from the Thompson scheme. This is particularly evident in areas coincident with regions typically associated with stratus in the EP basin (not shown).

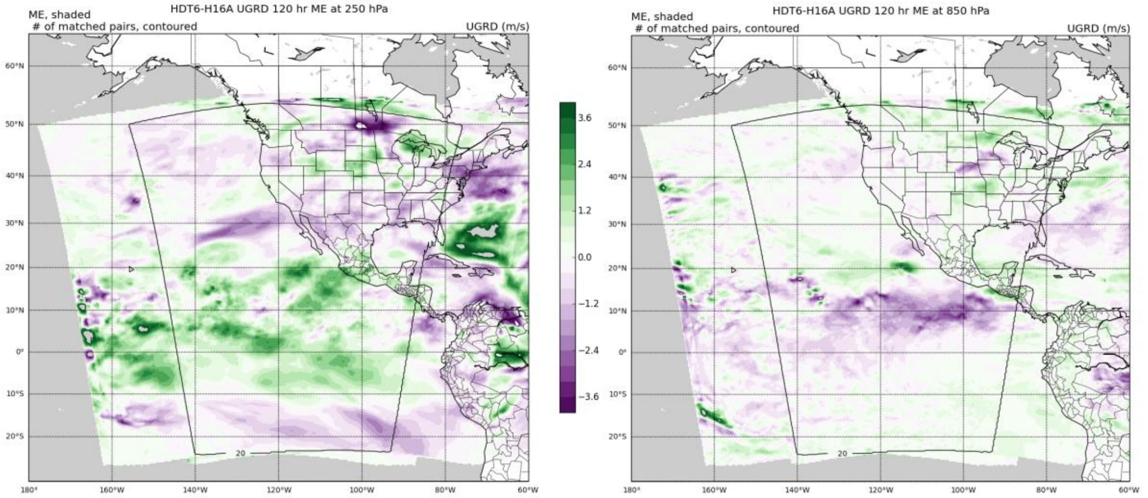


Figure 3.2.3-3. Large-scale verification of the zonal wind for aggregated 120 hour forecast lead times at 250 hPa (left) and 850 hPa (right). Shading indicates the difference of the mean errors of the Thompson configuration minus the operational Ferrier-Aligo configuration. Green shading indicates more westerly flow and purple shading indicates less westerly flow.

Further investigation of modifications that may lead to improvements in the representation of clouds and radiation are underway for inclusion in the project final report, which will be posted on the DTC website. Modifications include changes to the partial cloudiness scheme and non-local mixing in the PBL scheme. Case studies for both basins are being studied to assess whether the modifications address shortcomings in the EP basin identified in the larger T&E experiment. AL basin cases are also being included in this study to ascertain whether changes aimed at improvements in the EP basin affect the behavior in the AL basin.

3.3 Data Assimilation

One paper associated with past DA activities was accepted by *BAMS*:

Shao, H., J. Derber, X.-Y. Huang, M. Hu, K. Newman, D. Stark, M. Lueken, C. Zhou, L. Nance, Y.-H. Kuo, B. Brown, 2015: *Bridging Research to Operations Transitions: Status and Plans of Community GSI.* Bull. Amer. Meteor. Soc. doi:10.1175/BAMS-D-13-00245.1.

3.3.1 GSI-Hybrid System for HWRF

The DTC continued to investigate improving tropical cyclone (TC) intensity forecasts using regional ensembles in the GSI-hybrid DA system. The DTC set up a complete hybrid EnKF-GSI DA configuration for the HWRF system (Fig. 3.3.1-1) to conduct regional ensemble experiments, and then compared the results to the 2014 operational HWRF system. Following previous studies, the DTC continued to use Hurricane Irene (2011) as the test case. Results showed the 2014 operational configuration (Fig. 3.3.1-2, green) generated the best intensity scores at analysis time for Hurricane Irene. However, TC intensity biases increased rapidly (positive to negative, "spin-down") and stayed relatively large throughout the rest of the forecast. A similar spin-down issue was confirmed using the 2015 HWRF system, but with smaller biases. Removing the vortex initialization step from the operational workflow (cyan) increased the magnitude of the intensity biases. Two experiments using an HWRF ensemble (red and blue) generated smaller intensity biases for forecasts beyond 12-24 h. The two-way hybrid system (blue) was set up based on the GFS DA scheme, using the GSI deterministic analysis to re-center the ensemble members at each analysis time. The one-way hybrid system (red) skipped the re-centering step. This re-centering step reduced the ensemble spread for TC center locations and intensity. Consequently, the one-way hybrid system using regional ensembles performed better than the other systems for TC intensity forecasts beyond 12 h. Descriptions of the test set-up, experimental design, and results, as well as the previous DA balance study, are provided in the final report available at http://www.dtcenter.org/eval/data_assim/.

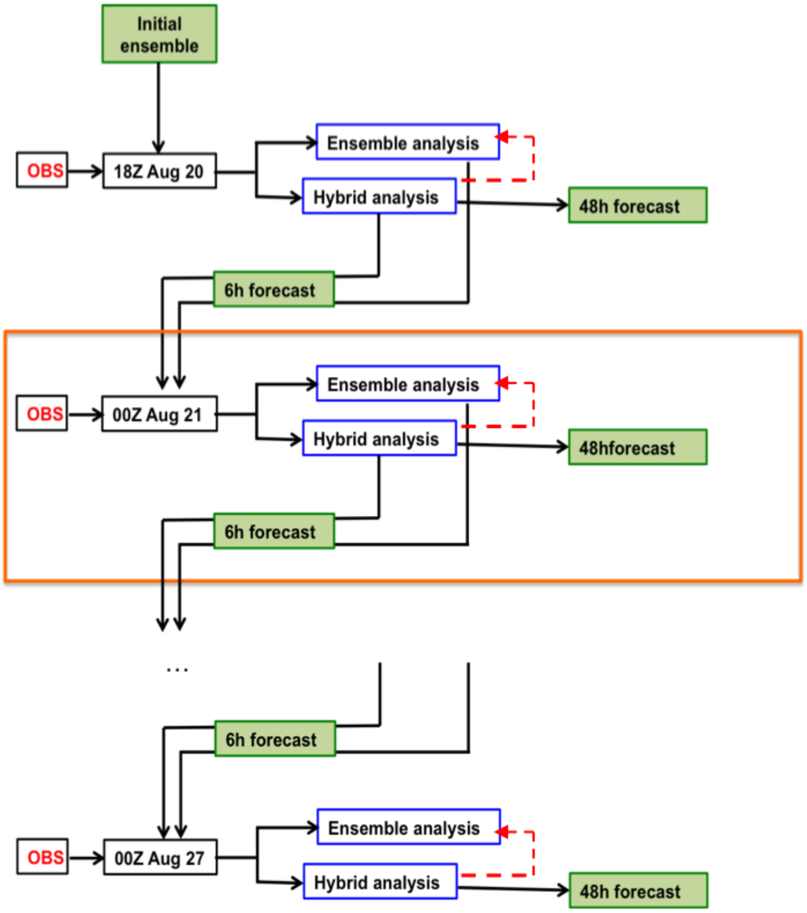


Figure 3.3.1-1. Schematic illustrating the hybrid EnKF-GSI DA procedure. Dashed line indicates the optional re-centering step in the hybrid system.

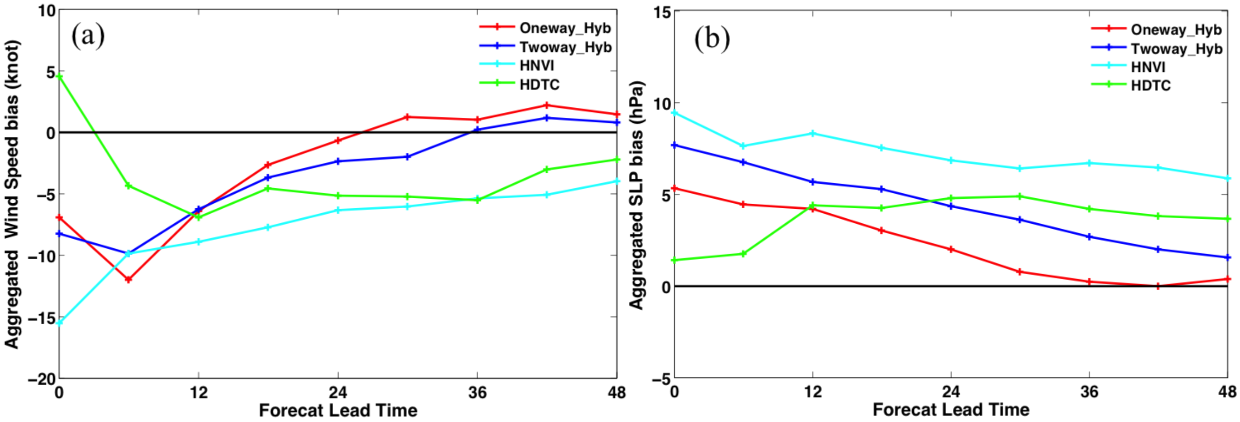


Figure 3.3.1-2. Aggregated bias for (a) maximum surface wind speed, and (b) minimum sea level pressure as a function of forecast lead time. Analyses and forecasts were generated using the HWRF system with the 2014 operational configuration (green) and without vortex initialization step using GFS ensemble (cyan), HWRF ensemble (red), and HWRF ensemble re-centered based on GFS deterministic analysis (blue).

3.3.2 Regional Ensemble Based DA T&E

The goals of this DA T&E activity were to examine the readiness of the GSI Ensemble-Variational (EnVar) system for regional applications and provide an assessment of potential areas for further improvement. The DTC produced a control run based on the NOAA RAP system, which currently uses the GSI 3D hybrid EnVar technique for its atmospheric DA. Then, an experimental end-to-end system, using the GSI 4D hybrid EnVar DA system, was built to investigate the EnVar capability and evaluate its impacts on analyses and forecasts.

As a first step, the DTC focused on case studies directed at exercising the GSI EnVar system to confirm it was configured properly. This testing phase included performing single observation tests, tuning the observation and ensemble/static error contributions for the hybrid option, and evaluating the merits of using regional ensembles for EnVar instead of using global ensembles (default for the operational RAP system). Figure 3.3.2-1 shows the temperature analysis increments from the pseudo single observation tests performed using the GSI EnVar system with the 3DVar, 3D hybrid EnVar, and 4D hybrid EnVar techniques, respectively. A single temperature observation was set at -3, 0 and +3 h within a 6-h time window. The results show that the analysis increments of the 4D EnVar experiments vary with the observation time, while the 3DVar and 3D hybrid experiments do not produce time-variant information. Both 3D and 4D EnVar experiments captured the flow-dependent features for the analysis fields due to the incorporation of ensemble-based background errors. These results, which are expected, demonstrated the GSI EnVar system is set up correctly and is ready for further testing. DTC staff also investigated the impact of replacing the 30-km GFS ensemble input (default RAP configuration) with 13-km RAP ensemble input, which reduced the RAP analysis errors, but had minimal impact on the forecasts. Consequently, the DTC decided to use GFS ensembles for its EnVar experiments.

To evaluate the feasibility of applying the 4D EnVar technique to an ARW-based regional system, the DTC ran three EnVar experiments for a two-week period with a simplified RAP framework (e.g., no non-variational cloudy analysis and digital filter prior to forecasts, 6-h cycling): 3DVar, 3D hybrid EnVar, and 4D hybrid EnVar. All three experiments used the observation files prepared for the GFS DA (e.g. through the quality control (QC) procedures specific for the GFS applications). Figure 3.3.2-2 shows RMSE of analyses and 6-h forecasts for humidity and wind fields. The 4D hybrid EnVar produced the smallest RMSE for most layers. The 4D experiment also achieved the smallest cost function values, showing improved convergence during the minimization process toward the final analysis. An assessment of the computational costs for these three experiments indicated one 4D hybrid EnVar run requires a 40% increase in computational wall-clock time over a single 3D hybrid EnVar run (using 384 processors on NCAR's Yellowstone computer).

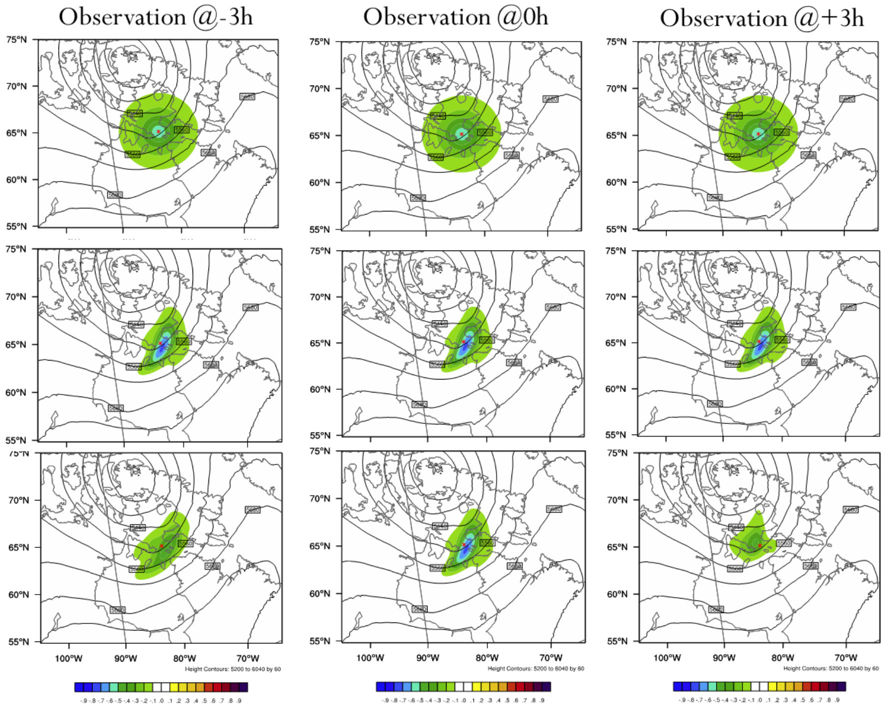


Figure 3.3.2-1. Temperature analysis increments from the pseudo-single observation tests for 3DVar (upper panels), 3D hybrid EnVar (middle panels), and 4D hybrid EnVar (lower panels). One single temperature observation was set at -3h (left column), 0 (middle column) and +3h (right column) within a 6h time window.

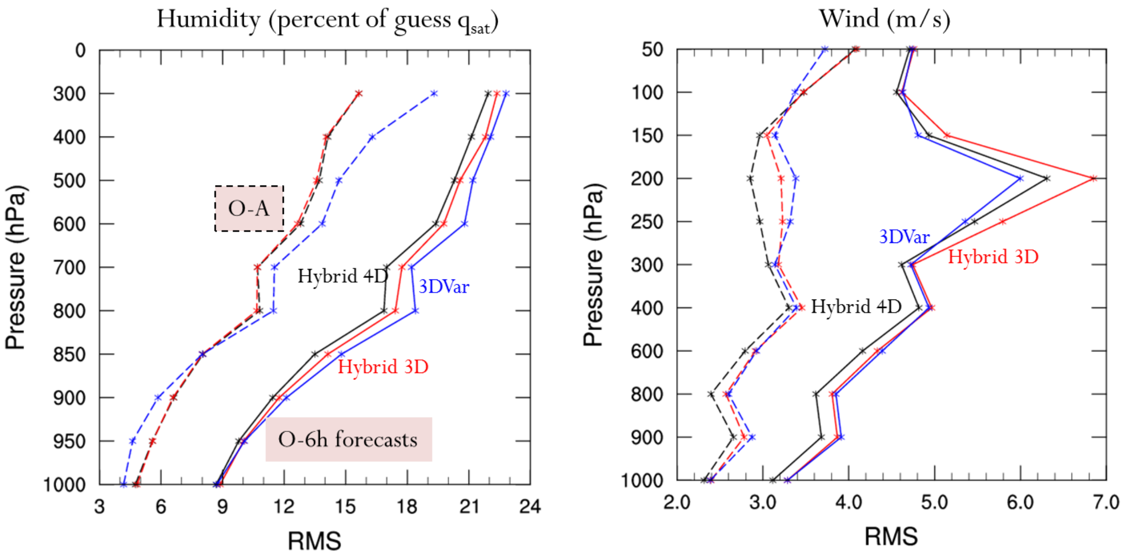


Figure 3.3.2-2. Vertical profiles of the RMSE of humidity (left) and wind (right) analyses (dashed lines) and 6h forecasts (solid) from three experiments: 3DVar (blue), 3D hybrid EnVar (red), and 4D hybrid EnVar (black).

The final stage of testing focused on the impact of 4D hybrid EnVar on the full blown hourly RAP system (i.e., including EnVar data assimilation, non-variational cloud analysis, and digital filter for the initialization step). The RAP-4D hybrid EnVar experiment generated larger biases and RMSE for temperature and wind than the RAP-3D hybrid EnVar experiment. Further study identified a number of problems with the current GSI system. For the 4D EnVar capability, the current GSI code does not correctly use the multiple time 2-m temperature background for the background array. This background field is used in RAP for surface DA, but did not affect the above-mentioned 6-h cycling experiments because the Global Data Assimilation System (GDAS) prepbufr files were used, which led to the surface observations being rejected (large QC flags) during the pre-processing step. The RAP-4D EnVar results were improved when the DTC corrected the multi-time level 2-m temperature background problem in the GSI code. The DTC committed this bug fix to the code repository, which was then shared with all developers, and recommended a thorough code review for all other background fields. The DTC also identified the current GSI EnVar is missing capabilities particular to ARW, including the dual resolution hybrid capability, the capability to use multiple-time level ARW ensembles for EnVar, and the capability to treat a GFS ensemble mean as one of the ensemble members in order to improve the ensemble representativeness on regional scales. All these experiments and results, as well as a summary of areas to further improve the current 4D EnVar capabilities for regional applications, will be described in a final report that will be posted on the DTC website.

3.4 Ensembles

3.4.1 Neural Network

Model uncertainty, and specifically uncertainty in physical parameterizations, is an important source of forecast error that can be represented through ensemble forecasting. Given complex microphysics schemes are computationally expensive, a neural network approach may provide stochasticity to ensembles with little additional computational time. To take initial steps toward this goal, the scope of the work conducted by the DTC was to produce datasets from NMMB simulations using a partial double-moment microphysics scheme, the Thompson parameterization. This activity was a carry-over activity from AOP 2014. To prepare the code, NMMB was run on the NOAA Research computational platforms, with testing to ensure bitwise identical simulations. This work was followed by implementation of write statements for critical microphysics variables just before and after calls to the Thompson microphysics scheme. Subsequently, a set of 16 simulations over the CONUS were produced using these added write statements. The datasets were provided to EMC for the purposes of training a neural network and testing it within NMMB.

3.4.2 North American Rapid Refresh Ensemble (NARRE) Repository and Rocoto End-to-end Workflow

For AOP 2015, the DTC established a functional workflow based on the Rocoto Workflow Management System for use by NCAR, EMC and GSD staff as a step towards NCEP's future HREF System. End-to-end workflows for RAP and NAMRR were established on Theia and Yellowstone, both of which included MET verification tasks for the deterministic forecasts and visualization scripts. A separate workflow was established to combine the deterministic RAP and NAMRR members using the Short-Range Ensemble Forecast post-processing (SREF-PP) package, followed by verification tasks that run MET on the ensemble output. A variety of ensemble products can be created by running the SREF-PP. The inclusion of MET in the workflow provides the opportunity to not only verify the final products from SREF-PP, but to also iteratively adjust the ensemble design while examining how probabilistic statistics change when different approaches are utilized. The NARRE data flow is show in Fig. 3.4.2-1

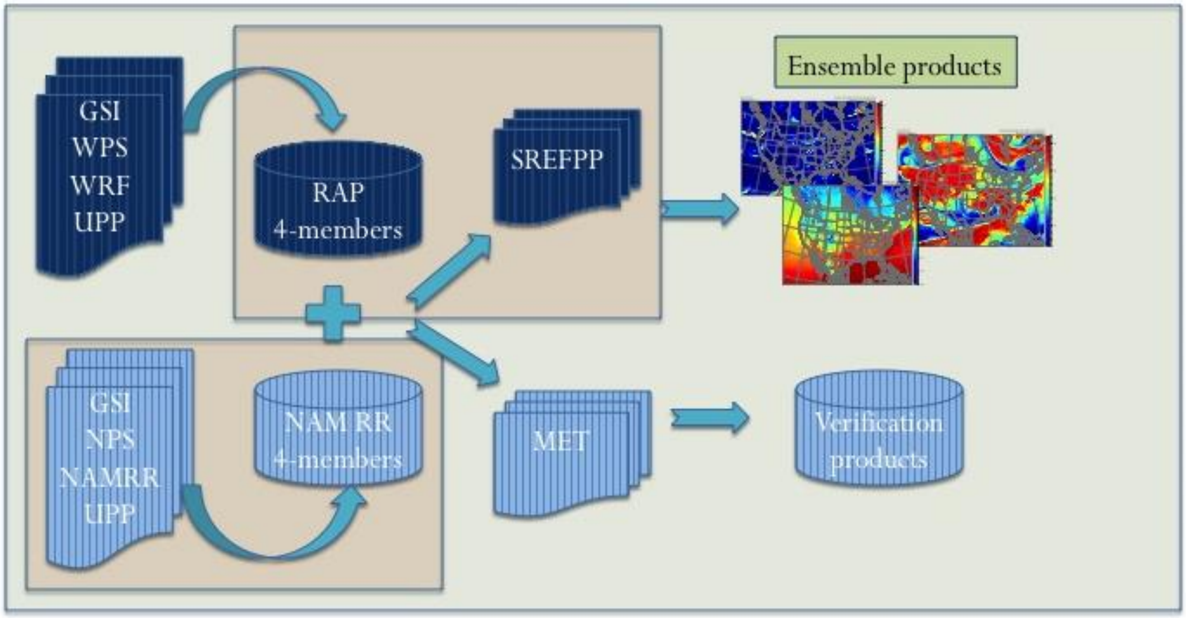


Figure 3.4.2-1. NARRE data flow.

3.4.3 Testing of stochastic physics for use in NARRE

Most global and regional NWP ensemble systems are under-dispersive, producing unreliable and overconfident ensemble forecasts. With growing evidence that initial-condition uncertainties are not sufficient to entirely explain forecast uncertainty, the role of model uncertainty is receiving increasing attention. In the last decade, a number of different strategies have been proposed to represent uncertainty arising from model error. These approaches include use of multi-dynamic core, multi-physics and a combination of multi-dynamic core and multi-physics. While the multi-physics approach yields desirable results, it has practical and theoretical deficiencies. Maintenance and development of a variety of physics is cost intensive. More importantly, this type of ensemble does not form a consistent distribution. In addition, each member has its own climatology and errors, which makes post-processing for these systems very challenging.

The focus of this activity was to compare the performance of the multi-physics and stochastic physics approaches. The operational RAP physics suite was used in the study. The stochastic approaches considered in this study were: i) stochastically perturbed parameter (Stoch), ii) stochastic kinetic energy backscatter (SKEB), and iii) stochastic perturbation of physics tendencies (SPPT). These approaches were combined to compose four stochastic configurations: 1) Stoch, 2) Stoch and SKEB (Stoch_skeb) 3) Stoch and SPPT (Stoch_sppt), and 3) Stoch, SKEB, and SPPT (Stoch_skep_sppt). The control for this experiment was a multi-physics configuration based on the current RAP members of the SREF. Parameter perturbations were introduced in the convective scheme (Grell-Freitas) for perturbations of closures and in the PBL scheme (Mellor-Yamada-Nakanishi-Niino - MYNN) for perturbations of mixing length, roughness length and cloud fraction.

Performance of the ensembles was evaluated in terms of bias, skill, accuracy, reliability and sharpness. The results confirm the findings of previous studies (e.g. Berner et al. 2011 and 2015 and Hacker et al. 2011). In particular, (1) parameter perturbations alone do not generate sufficient spread to remedy the under-dispersion in short-term ensemble forecasts and (2) a combination of several stochastic schemes outperforms any single scheme. This result implies that a synthesis of different approaches is best suited to capture model error in its full complexity.

Figure 3.4.3-1 shows the Continuous Rank Probability Score (CRPS) as one example of superior performance of the ensemble that combines several different stochastic approaches. This score is oriented such that a lower value denotes better forecast skill. Stoch is, for the most part, characterized by the highest CRPS (lowest skill) compared to the other ensembles. The one exception is the 6-h lead time of the 00 UTC initialization for 500-hPa geopotential height. For most of the variables and most lead times, the ensemble that combines three stochastic approaches is characterized by the best forecast skill, even though differences in CRPS between spp_skeb_sppt and the control ensembles were not statistically significant.

This activity's findings may potentially have a large impact on the design of next-generation high-resolution regional, as well as global, operational ensembles. The single-physics stochastic approach clearly provides a viable alternative to using the costly, and theoretically problematic, multi-physics approach in operations. More details about this study are provided in the project report on the DTC website http://www.dtcenter.org/eval/ensembles/Report_EN2_2015.pdf.

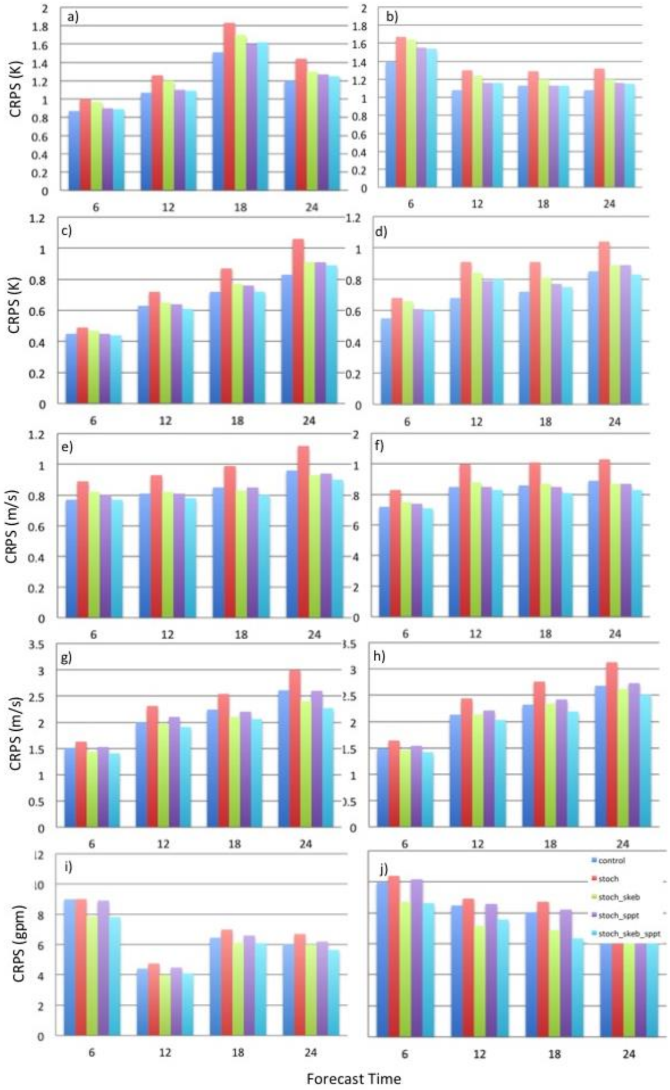


Figure. 3.4.3-1. CRPS for 2-m temperature (a and b), 850-hPa temperature (c and d), 10-m zonal wind component (e and f), 250-hPa zonal wind component (g and h), and 500-hPa geopotential height (i and j), for 00 UTC initializations (left column) and 12 UTC initializations (right column). The color bars denote the five experiments described in the text.

3.5 Global Model Test Bed (GMTB)

To facilitate the development of an advanced physics suite for NWS's NGGPS, the DTC is developing a uniform 'test harness" to enable in-depth investigation of various physical parameterizations. The test harness mimics the logical progression for testing newly developed parameterizations that typically takes place within the scientific community. Components are gradually added as one moves through the hierarchy until the full forecast model complexity is reached. It is designed to complement both the existing testing protocol at EMC and independent testing typically performed by parameterization developers. Figure 3.5-1 illustrates the hierarchical tiers of the test harness and represents how the DTC envisions the division of effort (GMTB's likely role denoted by blue) and how the harness fits within EMC's existing testing framework.

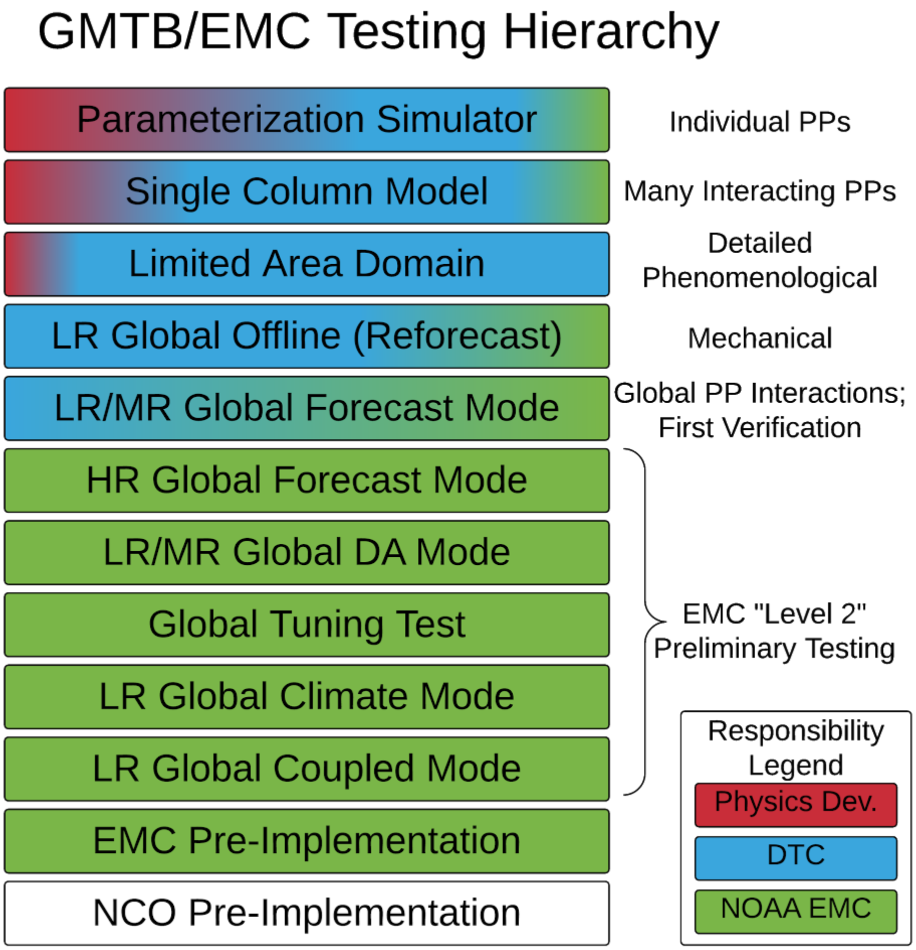


Figure 3.5-1. Diagram illustrating the testing hierarchy plan to support physics development for NGGPS. LR indicates low resolution, MR medium resolution, and HR high resolution. Color shading indicates where the different groups are anticipated to focus their efforts (red – physics developers, blue – GMTB task within the DTC, and green – EMC). PP stands for physics parameterization.

3.5.1 Single Column Model (SCM)

To fulfill GMTB's goals, any SCM used within the testing framework must utilize the IPD and be minimally tied to existing global models. This approach ensures compatibility with both physical parameterizations that may comprise the CCPP now or in the future and the new dycore selected for NGGPS through support of multiple vertical coordinates. Since existing SCM code at EMC does not meet

GMTB’s objectives, development of a new SCM at the DTC is underway. In addition to meeting the design criteria mentioned above, the new code will be “community-friendly.” It uses modern Fortran90 coding standards and netCDF standard for I/O that can be compiled using the freely-available GFortran compiler. Python scripts for basic plotting and thorough documentation using the Doxygen tool are under development.

SCMs are typically driven in one of two ways: through idealized initial conditions and forcing derived from intensive observational period data or via global forecast model output. GMTB is focusing its initial efforts on the former method using a collection of cases created by the Global Energy and Water cycle EXchanges (GEWEX) Global Atmospheric System Studies (GASS) and European Union Cloud Intercomparison, Process Study and Evaluation (EUCLIPSE) programs, but support for the second method of driving the SCM will follow once a collection of idealized cases is established. The cases represent a broad range of meteorological conditions in geographic locations around the globe such that testing a physics suite using the entire collection will provide a quick overview of how the suite might behave in a global, three-dimensional forecast. For prototype testing purposes, a Lagrangian stratocumulus-to-cumulus transition case based on the Atlantic Stratocumulus to cumulus Transition EXperiment (ASTEX) field campaign is being used for two reasons. First, a recent inter-comparison study (van der Dussen, 2013, JAMES) using the case demonstrated that the regime transition is challenging for physics schemes to accurately represent. Second, although some cases specify that only a subset of a physics suite should be active (a configuration somewhere between a “parameterization simulator” and a fully interactive SCM in the testing hierarchy), this case features full interactivity of all parameterization types within a suite.

3.5.2 Workflow for LR/MR Global Forecast Mode

To facilitate three-dimensional testing that provides information about the interaction between the physics packages and feedback on the large-scale flow, the GMTB is assembling an end-to-end workflow for GFS free forecasts on NOAA’s research computer Theia. This workflow is heavily based on the scripts EMC uses to produce its parallel NEMS/GSM runs. The EMC capabilities that will be supported by the GMTB include the scripts to run the GSM and UPP, which is complete. Current efforts are focused on making the workflow more flexible to meet GMTB’s needs. The GMTB has established a collaborative dialogue with EMC’s Global Team regarding the migration of their current scripting architecture to a unified global workflow. This effort includes GMTB members attending weekly meetings that EMC has instituted to discuss the direction and progress of the unified global workflow.

As an avenue to facilitate development and in-depth evaluation of physics parametrizations and physics suites, the GMTB compiled a list of diagnostics and comprehensive verification tools to help support the developers. An initial workflow was established to create basic diagnostics with Python scripts (e.g., temperature, moisture, wind, and height fields at different levels) and run MET to provide verification metrics. The verification portion of the workflow is being configured to replicate key functionality of EMC’s verification package (e.g., near-surface, upper-air, and precipitation verification) as well as provide additional, more advanced verification metrics (e.g., fractions skill score). The diagnostics and verification workflow has been designed with flexibility and portability in mind to allow for ease of use by the model physics community. Work is currently underway to expand the capabilities in the testbed in order to equip physics developers with a wide spectrum of tools to assess strengths and deficiencies of physics parameterizations. For example, the GMTB is looking to include the calculation of time-integrated tendencies, sub-grid fluxes, cloud properties, and conservation of select fields.

To manage the workflow and scripts that are being developed and used by the GMTB, a Git repository has been set-up through Vlab. The repository contains the Python scripts for creating plots of diagnostics and the scripts and configuration files for running MET.

3.5.3 Sea Ice Model Test

Given the recommendation for NGGPS to adopt the Los Alamos Sea Ice model (CICE; Hunke and Lipscomb 2008) as a component of UGCS, GMTB developed a plan with input from EMC to carry out preliminary T&E of CICE. The test will be conducted in two phases, both involving the CICE model run in standalone mode, forced by prescribed atmospheric and oceanic fields. The test design took into consideration the results by Hebert et al. (2015), who showed that the Arctic Cap Nowcast/Forecast System (ACNFS) demonstrated a high level of skill compared to persistence in 1-7 day forecasts over a period of one year. ACNFS uses CICE version 4.0 (Hunke and Lipscomb 2008) as the sea ice model, two-way coupled to the HYbrid Coordinate Ocean Model (HYCOM - Bleck, 2002; Metzger et al., 2014, 2015), both forced by the Navy Global Environmental Model (NAVGEM) atmospheric fields.

In order to make the GMTB test most relevant for NGGPS, CICE version 5, which includes a number of new physics options, such as the mushy-layer thermodynamics and two new melt pond parameterizations, will be employed. In addition, the GMTB test will use atmospheric fields from NCEP's global modeling applications to force the ice and ocean fields. It is anticipated that the most challenging aspect of this test will be creating good quality sea ice initial conditions. Climate Forecast System version 2 (CFSv2) will provide ice/snow temperatures, concentration, thickness and velocity on four layers. A methodology to initialize CICE's seven layers will need to be devised before starting the test.

Given the NGGPS's goal of providing improved accuracy for forecasting from a few hours to a month for resolutions ranging from 1 to 100 km, it is anticipated that options within the sea ice model will likely be necessary to meet the needs of the various NGGPS forecast applications. Thus, the DTC will test the model for two different horizontal resolutions: Phase 1 will use CICE configured with 100-km horizontal grid spacing, while in Phase 2 CICE will use 15-km grid spacing (see Table 3.5.3-1)

Table 3.5.3-1. Configuration for Phases 1 and 2 of the CICE test.

	Phase 1	Phase 2
CICE resolution	100 km	15 km
Initial conditions	CFSv2 Reanalysis	CFSv2 Reanalysis
Atmospheric forcing	CFSv2 Reanalysis	GFS ¼ degree
Ocean forcing	CFSv2 Reanalysis	CFSv2 Reanalysis
Domain	Global	
Forecast length	16 days	
Number of cases	24	
Initialization	1 and 15th of every month for 2015	

Sea ice forecast verification against the NCEP 1/12th degree analyses and observations will be performed by GMTB and EMC collaborators.

4 References

Bleck, R., 2002: An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates, *Ocean Modelling*, **4(2)**, 219–219, doi: 10.1016/S1463-5003(01)00017-8.

Berner, J., S.-Y. Ha, J. P. Hacker, A. Fournier, and C. Snyder, 2011: Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multi-physics representations. *Mon. Wea. Rev.*, **139**, 1972–1995.

Berner, J., K. R. Fossell, S.-Y. Ha, J. P. Hacker, and C. Snyder, 2015: Increasing the skill of probabilistic forecasts: Understanding performance improvements from model-error representations. *Mon. Wea. Rev.*, **143**, 1295-1320, doi:10.1175/MWR-D-14-00091.1.

Hacker, J. P., C. Snyder, S.-Y. Ha, and M. Pocernich, 2011: Linear and nonlinear response to parameter variations in a mesoscale model. *Tellus A*, 63, 429–444.

Hebert, D. A., R. A. Allard, E. J. Metzger, P. G. Posey, R. H. Preller, A. J. Wallcraft, M. W. Phelps, and O. M. Smedstad, 2015: Short-term sea ice forecasting: An assessment of ice concentration and ice drift forecasts using the U.S. Navy's Arctic Cap Nowcast/Forecast System, *J. Geophys. Res. Oceans*, **120**, 8327-8345, doi:10.1002/2015JC011283.

Hunke, E., and W. Lipscomb, 2008: CICE: The Los Alamos Sea Ice Model documentation and software user's manual version 4.0, Tech. Rep. LA-CC-06-012, Los Alamos Natl. Lab., Los Alamos, N. M.

Kaplan, J. and M. DeMaria, 2003: Large-scale characteristics of rapidly intensifying tropical cyclones in the North Atlantic basin. *Wea Forecasting*, **18**, 1093- 1108.

Metzger, E.J., O.M. Smedstad, P.G. Thoppil, H.E. Hurlburt, J.A. Cummings, A.J. Wallcraft, L. Zamudio, D.S. Franklin, P.G. Posey, M.W. Phelps, P.J. Hogan, F.L. Bub, and C.J. DeHaan, 2014: US Navy operational global ocean and Arctic ice prediction systems, *Oceanography*, **27(3)**, 32–43, http://dx.doi.org/10.5670/oceanog.2014.66.

Metzger, E. J., H. E. Hurlburt, A. J. Wallcraft, J. F. Shriver, O. M. Smedstad, P. Thoppil, and D. S. Franklin, 2008: Validation test report for the Global Ocean Prediction System v3.0 - 1/128 HYCOM/NCODA phase 1. Tech. Rep. NRL/MR/7320-0809148, Naval Res. Lab., Stennis Space Cent., Hancock, Miss.

van der Dussen, J. J., S. R. de Roode, A. S. Ackerman, P. N. Blossey, C. S. Bretherton, M. J. Kurowski, A. P. Lock, R. A. J. Neggers, I. Sandu, and A. P. Siebesma, 2013: The GASS/EUCLIPSE model intercomparison of the stratocumulus transition as observed during ASTEX: LES RESULTS*. Journal of Advances in Modeling Earth Systems*, **5(3),** 483–499.

5 Acronyms and Abbreviations

ACARS	Aircraft Communications Addressing and Reporting System
ACNFS	Arctic Cap Nowcast/Forecast System
AERONET	AErosol Robotic NETwork
AF	Air Force
AL	Atlantic basin
AMS	American Meteorological Society
AOML	Atlantic Oceanographic and Meteorological Laboratory
AOP	Annual Operating Plan
AR	Atmospheric River
ARW	Advanced Research WRF
ASTEX	Atlantic Stratocumulus to cumulus Transition EXperiment
BAMS	Bulletin of the American Meteorological Society
BCRMSE	Bias-Corrected Root Mean Squared Error
CCPP	Common Community Physics Package
CFSv2	Climate Forecast System version 2
CI	Confidence Interval
CICE	Los Alamos Sea Ice Model
CMORPH	CPC MORPHing technique
CONUS	Contiguous United States
CRPS	Continuous Rank Probability Score
DA	Data Assimilation

DTC	Developmental Testbed Center
EC	Executive Committee
EMC	Environmental Modeling Center
EnKF	Ensemble Kalman Filter
EnVar	Ensemble-Variational
EP	Eastern North Pacific
ESRL	Earth System Research Laboratory
EUCLIPSE	European Union Cloud Intercomparison, Process Study and Evaluation
FAR	False Alarm Ratio
GASS	Global Atmospheric System Studies
GCWMB	Global Weather and Climate Modeling Branch
GDAS	Global Data Assimilation System
GEWEX	Global Energy and Water cycle EXchanges
GFDL	Geophysical Fluid Dynamics Laboratory
GFS	Global Forecast System
GMAO	Global Modeling and Assimilation Office
GMTB	Global Model Test Bed
GOES	Geostationary Operational Environmental Satellite
GOSA	Global Observing Systems Analysis
GSD	Global Systems Division
GSI	Gridpoint Statistical Interpolation
GSS	Gilbert Skill Score
H214	2014 HWRF Retrospective Test
HFIP	Hurricane Forecast Improvement Project
HIWPP	High Impact Weather Prediction Program
HMT	Hydrometeorology Testbed
HR	High Resolution
HRD	Hurricane Research Division
HREF	High Resolution Ensemble Forecast
HRRR	High Resolution Rapid Refresh
HWRF	Hurricane WRF
HYCOM	HYbrid Coordinate Ocean Model
IPD	Interoperable Physics Driver
JCSDA	Joint Center for Satellite Data Assimilation
LW	Longwave
LR	Low Resolution
MAE	Mean Absolute Error
MB	Management Board
MCC	Mesoscale Convective Complexes
MET	Model Evaluation Tools
MMB	Mesoscale Modeling Branch
MMET	Mesoscale Model Evaluation Testbed
MMM	Mesoscale and Microscale Meteorology (Laboratory at NCAR)
MODE	Method for Object-based Diagnostic Evaluation
MPIPOM-TC	Message Passing Interface Princeton Ocean Model for Tropical Cyclones
MR	Medium Resolution
MTD	MODE Time Domain
MYNN	Mellor-Yamada-Nakanishi-Niino

NAM	North American Mesoscale
NAMRR	NAM Rapid Refresh
NARRE	North American Rapid Refresh Ensemble
NASA	National Aeronautics and Space Administration
NAVGEM	Navy Global Environmental Model
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
NCL	NCAR Command Language
NCO	NCEP Central Operations
NCWCP	NOAA Center for Weather and Climate Prediction
NEMS	NOAA Environmental Modeling System
NESDIS	National Environmental Satellite, Data and Information Service
NetCDF	Network Common Data Form
NGGPS	Next Generation Global Prediction System
NHC	National Hurricane Center
NITE	NWP Information Technology Environment
NMMB	Nonhydrostatic Multiscale Model on the B grid
NMME	Nonhydrostatic Mesoscale Model on the E grid
NOAA	National Oceanic and Atmospheric Administration
NPS	NMMB Preprocessing System
NRL	Naval Research Laboratory
NSF	National Science Foundation
NUOPC	National Unified Operational Prediction Capability
NWP	Numerical Weather Prediction
NWS	National Weather Service
OAR	Office of Oceanic and Atmospheric Research
ONR	Office of Naval Research
PBL	Planetary Boundary Layer
PHIST	Probability Integral Transform Histogram
POD	Probability of Detection
PP	Physics Parameterizations
PS	Practical Significance
PSD	Physical Sciences Division
QC	Quality Control
QPE	Quantitative Precipitation Estimate
QPF	Quantitative Precipitation Forecast
R2O	Research to Operations
R&D	Research and Development
RAMADDA	Repository for Archiving, Managing and Accessing Diverse DAta
RAP	Rapid Refresh
RI	Rapid Intensification
RMSE	Root Mean Square Error
RRTM	Rapid Radiative Transfer Model
RRTMG	Rapid Radiative Transfer Model for General Circulation Models
RT	Request Tracker
SAB	Science Advisory Board
SCM	Single Column Model
SKEB	Stochastic Kinetic Energy Backscatter

SPPT	Stochastic Perturbation of Physics Tendencies
SREF	Short-Range Ensemble Forecast
SREF-PP	Short-Range Ensemble Forecast Post-Processing
SS	Statistical significance
SUNY	State University of New York
SW	Shortwave
T&E	Testing and Evaluation
TC	Tropical Cyclone
TOO	Terms of Operation
UGCS	Unified Global Coupled System
UPP	Unified Post-Processor
URI	University of Rhode Island
USWRP	US Weather Research Program
VAPOR	Visualization and Analysis Platform for Ocean, Atmosphere and Solar Researchers
WAF	Weather Analysis and Forecasting
WP	Western North Pacific
WPC	Weather Prediction Center
WRF	Weather Research and Forecasting
WWSIS	Western Wind and Solar Integration Study